



Diseño de Experimentos Dinámicos basado en Modelos para Optimización

Caso de Estudio: Producción en cultivo de alta densidad celular del antígeno
rBm95 para vacuna contra *Rhipicepalus (Boophilus) microplus*

Tesis Doctoral para optar por el título de Doctor en Tecnología Química
Facultad de Ingeniería Química
Universidad Nacional del Litoral

Autor: Mariano D. Cristaldi

Director: Ricardo J. A. Grau

Co-director: Ernesto C. Martínez

Jurados:

Dr. Carlos E. Argaraña, CIQUIBIC (CCT Córdoba – UNC)

Dr. José A. Bandoni, PLAPIQUI (CCT Bahía Blanca – UNS)

Dr. Ricardo B. Kratje, Instituto de Tecnología Biológica (UNL)

Santa Fe, 2011

*Buenas noches familiares y amigos.
Gracias por estar aquí.
Espero que disfruten de est espectáculo de luz y de sonido.
Esta noche trataré de que nos sintamos menos solos, trataré de mostrar que estamos vivos y que no todo
está escrito.
En 1927 un matemático formuló el principio de incertidumbre.
Venía a decir algo así como que nada se puede predecir con exactitud,
siempre queda un margen de incertidumbre en el conocimiento humano, y en ese margen de
incertidumbre yo siempre pensé que estaba la música, las canciones (el amor).
El principio está relacionado con el hecho de que el observador, con el mero hecho de ser testigo,
influye en la realidad que está observando, la altera, introduce una variable de indeterminación,
y esta noche, si a ustedes les parece bien, me gustaría hacer un experimento.
Me gustaría mostrar que cada canción es diferente simplemente porque tú estás a mi lado,
cada concierto es diferente porque tú lo escuchas, porque tú cantas conmigo.
Así que, manos a la obra. Nada está escrito, la historia no ha terminado.
Quizá los siguientes días sigan siendo terribles y grises, puede ser,
pero puede que no, puede que todo cambie,
que los días que tienen que venir abran ventanas a la esperanza.
Este puede ser un buen comienzo, este puede ser un buen principio.*

*Ismael Serrano
Apertura de Concierto Principio de Incertidumbre*

GRACIAS a la vida, que me ha dado **TAN**TO
Me dio dos luceros, que cu**AN**do los abro
Perfecto distingo, lo negro del bl**AN**co
y en el al**TO** ci**EL**o, su fondo estrell**AD**o
y en las multitudes el **HombrE** que yo amo. g**R**acias a la vida, que **Me** ha dado t**AN**to me h**A(S)**
dado el oído que en todo **Su anchO** gra**Ba** noche y día, g**R**illos y ca**N**ario**S**
martillos, turbinas, ladridos, chubascos
y la voz tan tiern**A** de **MI** bien amado. Gracias a la vida, que me ha dad**O** tanto**(S)**
me a dado el so**NI**do y el abe**CedariO**
con él las palabras que pien**SO** y declaro ma**Dre**, Amigo, **Hermano** y luz
a**L**umbran**DO**
la ruta del al**MA** del que estoy amando. g**R**acias a la vida, que me ha dado Tanto**(U)**
me a dado la ma**R**cha de m**IS** pies cansado con ellos anduve ciudades y **CHAR**cos playas y **Desiertos**,
montañas y llanos
y en la casa tuya,
tu calle y tu patio. g**R**acias a la vida, que me ha dado ta**N**tom**E** dio el corazón que agita **Su** marco cuando miro el
fru**TO** del cerebro humano
cuando **M(a)I**ro el bue**N**o tan **I**ej**S** del malo
cuando miro el fondo de tus o**J**os claros**(E)**. gracia**S** a la v**IDA**, que me ha dado tanto
Me ha dado l**A** **R**isa y me ha dado el llanto
así yo distingo dicha de que**R**anto los dos mater**IA**les que forman mi can**T**o y el c**AN**to de ustedes que es el mismo
canto
y el c**AN**to de **TODOS** que es mi propio canto

GRACIAS a la vida...

Agradecimientos

En primer lugar quiero agradecer a mi familia sanguínea ya que sin ellos no hay magia posible: Mamá, Papá, Florencia, Cecilia, Clidia, Punchi, Martina, Rita, Hugo, Emiliano...gracias por hacerme feliz y sentir que soy parte de algo más grande, que nada es tan trágico ni malo como para no poder sonreír, por darme tantos ejemplos prácticos de que la única lucha que se pierde es la que se abandona...por tanta resiliencia.

Ahora sí! Empieza la familia ampliada...

Al Tato!...gracias por las peleas, los abrazos, las risas y la edición de los gráficos.

Al Nico! gracias por tanta compañía, oído, pragmatismo, buena comida, asistencia técnica e informática, mates, tererés y tantas cosas.

Richard (Dr. Ricardo Grau)!...gracias Grau por ser más que un director...por los festejos, las oportunidades, por las peleas, el acompañamiento en general pero particularmente en los fracasos y momentos críticos que fueron muchos y de variada intensidad y naturaleza...espero haber satisfecho, al menos en parte, lo que se esperaba de mí y algún día poder recompensarte.

María Inés (Dra. María Inés Cabrera) gracias por toda la ayuda recibida en estos años juntos y particularmente en la corrección de esta tesis, por la paciencia, la atención y la solidaridad intelectual.

Ernesto (Dr. Ernesto Martínez)! gracias por nuestras sendas charlas que discurrieron entre filosofía, pragmatismo, idealismo, realismo...y andá a saber cuántas cosas más! Por todas las manos que ofreciste y la camaradería.

Gracias a mis amigos entrañables Martín, Sonia, Dani y Aye por tantos momentos inolvidables y perdón por todos los olvidables que seguramente son muchos.

Gracias a mis amigos y compañeros de Química Fina, por su conocimiento, compañía, contención, consuelo, solidaridad, y tantos etcéteras

A mis compañeros de militancia, de circo, de acrobacias...gracias por hacerme ver que afuera hay un mundo diverso y múltiple que no siempre es lineal y que puede leerse e interpretarse en más de un registro.

Agradezco a José Di Conza por la gran mano en hacer realidad la posibilidad del trabajo experimental en microbiología y biología molecular.

Gracias a la Cátedra de Microbiología General de la FBCB-UNL por permitirme realizar mis actividades en sus instalaciones y ser tan amables conmigo, especialmente a

la Lic. Ayelén Porto por su colaboración, enseñanza, transferencia de *expertise* y tanta solidaridad.

Gracias al Dr. Adrián Filiberti del CIQUIBIC-UNC por su ayuda en biología molecular.

Gracias al INTEC-UNL y a la FIQ-UNL.

Gracias al Consejo Nacional de Investigaciones Científicas y Técnicas y a cada uno de los argentinos siguen apostando a que una sociedad próspera es una basada en el conocimiento. Gracias por hacer posible la realización de este trabajo de tesis aportando los recursos económicos necesarios.

No merezco tanto...

GRACIAS DIOS

*Mariano Cristaldi
febrero de 2011*

Los resultados obtenidos durante la realización de esta tesis han dado origen a los siguientes trabajos:

Publicaciones en revistas internacionales con referato:

Autores: Ernesto C. Martínez, **Mariano D. Cristaldi**, Ricardo J.A. Grau
“Design of Dynamic Experiments in Modeling for Optimization of Batch Processes”
Industrial & Engineering Chemistry Research, Vol. **48**(7), 3453-3465 (2009). ISSN 0888-5885.

Autores: **Mariano D. Cristaldi**, Ricardo J.A. Grau, Ernesto C. Martínez
“Iterative Design of Dynamic Experiments in Modeling for Optimization of Innovative Bioprocesses”
Process Modeling & Control-A Special Issue of Chemical Product & Process Modeling, Vol **4**, Issue 2, Article 6 (2009). ISSN: 1934-2659.

Autores: **Mariano D. Cristaldi**, María I. Cabrera, Ernesto C. Martínez, Ricardo J.A. Grau
“Finding the Simplest Mechanistic Kinetic Model Describing the Homogeneous Catalytic Hydrogenation of Avermectin to Ivermectin”
Industrial & Engineering Chemistry Research, ie-2010-01289h. ISSN: 0888-5885. (aceptado).

Autores: **Mariano D. Cristaldi**, Ernesto C. Martínez, Ricardo J.A. Grau
“A Model-based Policy Iteration Approach for run-to-run Optimization of Bioprocess Productivity”
Biotechnology and Bioengineering (en proceso de escritura, 2011).

Autores: **Mariano D. Cristaldi**, Ernesto C. Martínez, Ricardo J.A. Grau
“Optimization of rBm95 production in a *Escherichia coli* culture using a probabilistic tendency model based algorithm for experimental design”
Biotechnology and Bioengineering (en proceso de escritura, 2011).

Congresos Internacionales:

Autores: Ernesto C. Martínez, **Mariano D. Cristaldi**, Ricardo J.A. Grau, Joao Lopes
“Dynamic optimization of bioreactors using probabilistic tendency models and bayesian active learning”
European Symposium on Computer-Aided Process Engineering. ESCAPE 21
Chalkidiki, Greece.
29 Mayo – 1 Junio de 2011.
(en evaluación)

Autores: **Mariano D. Cristaldi**, Smaranda Cristea, Ernesto C. Martínez
“Run-to-run Convergence Analysis of Model based Policy Iteration”
European Symposium on Computer-Aided Process Engineering. ESCAPE 20
Ischia, Nápoles, Italia.
6-9 Junio de 2010.
pp. 925-930. ISBN: 978-044-453-569-6

Autores: **Mariano D. Cristaldi**, Ernesto C. Martínez, Ricardo J. A. Grau, María I. Cabrera.

“Hidrogenación de Avermectinas en Medio Homogéneo. Discriminación de Modelos Cinéticos Utilizando Análisis de Sensibilidad Global”

9° Congreso Interamericano de Computación Aplicada a la Industria de Procesos. CAIP 2009.

Montevideo, Uruguay.

25-28 Agosto de 2009.

pp. 395-400. ISBN: 978-956-319-925-3.

Autores: **Mariano D. Cristaldi**, Ricardo J.A. Grau, Ernesto C. Martínez

“Optimización Iterativa de Bioprocesos utilizando Diseño de Experimentos Dinámicos Basado en Modelos de Tendencia”

9° Congreso Interamericano de Computación Aplicada a la Industria de Procesos. CAIP 2009.

Montevideo, Uruguay.

25-28 Agosto de 2009.

pp. 389-394. ISBN: 978-956-319-925-3.

Autores: **Mariano D. Cristaldi**, Ernesto C. Martínez, Ricardo J.A. Grau

“Sequential Design of Dynamic Experiments in Modeling for Optimization of Biological Processes”

Abstract 28-4.

Sección: Modeling, Supporting Methods and Tools.

10th International Symposium on Process Systems Engineering - PSE2009.

San Salvador de Bahía, Brasil.

16-20 Agosto de 2009.

Publicado en *Computer Aided Chemical Engineering*, Vol 27, 369-374 (2009).

Autores: **Mariano D. Cristaldi**, Ernesto C. Martínez, Ricardo J.A. Grau

“Model-Based Iterative Design of Dynamic Experiments in Modeling for Optimization of a Bioprocess for a Recombinant Protein Production”

CHEMPOR 2008.

10 th International Chemical and Biological Engineering Conference.

Primary Area: Systematic Methods and Tools for Managing the Complexity.

Paper 112.

ISBN: 978-972-97810-3-2.

Braga, Portugal.

4-6 Septiembre de 2008.

Índice General

RESUMEN	vii
INTRODUCCIÓN	ix
Bibliografía	xiv
Capítulo 1: CONCEPTOS GENERALES	
1.1. Modelos	2
1.1.1. Concepto	2
1.1.2. Estructura y parámetros. Incertidumbre estructural y paramétrica	2
1.1.3. Clasificación	4
1.2. Análisis de Sensibilidad de Modelos	5
1.2.1. Generalidades	5
1.2.2. Propiedades de un AS ideal	6
1.2.3. Implementación de AS	6
1.2.4. AS local y global	7
1.3. Priorización de Factores (<i>FP</i>)	8
1.4. Mapeo de Factores. Monte Carlo Filtering	9
1.5. <i>GSA</i> en la práctica	10
1.5.1. Ejemplo de aplicación: Modelado de hidrogenación de avermectinas en medio homogéneo utilizando el catalizador de Wilkinson	10
1.5.2. Metodología propuesta para el modelado de reacciones químicas	10
1.6. Conclusiones	13
1.7. Bibliografía	14
Capítulo 2: MODELADO PARA OPTIMIZACIÓN	
2.1 Hipótesis de Trabajo	17
2.1.1. Modelos de Tendencias. Concepto	17
2.1.2. Modelado para Optimización. Concepto	17
2.1.3. Errores estructural, paramétrico y total de los modelos	19
<i>Error paramétrico</i>	21
<i>Error estructural</i>	22
<i>Error total del modelo</i>	22
Criterio de selección del “mejor” modelo	23
2.2. Diseño de Experimentos Dinámicos para Optimización basado en Modelos	25
2.2.1. Metodología propuesta	25
2.2.1.1. Iteración de la política de operación basada en modelos	26
2.2.1.2. Muestreo óptimo	29
2.2.2. Actualización óptima de modelos y selección del mejor modelo para optimización	30
2.2.3. Convergencia del método	31
Condiciones para la convergencia	33
Definición 1. El sumidero de Θ	34
Definición 2. Descenso suficiente del algoritmo de identificación-optimización	34
Propuesta	35
2.3. Ejemplo 1: Diseño de Experimentos Dinámicos para la Optimización de la producción de penicilina en un cultivo de <i>Penicillium crysogenum</i> . Proceso idéntico al modelo	36
Caso 1: Perfil de alimentación constante	39
Caso 2: Perfil de alimentación siguiendo un polinomio inverso de 1er orden	41
Caso 3: Perfil de alimentación siguiendo un polinomio inverso de 2do orden	44
Caso 4: Perfil de alimentación del biorreactor siguiendo un polinomio inverso de 3er orden	49
2.3.2 Conclusiones	53
2.4.1 Ejemplo 2: Optimización iterativa de la producción de la hormona peptídica interleukina-11 recombinante en <i>Escherichia coli</i>	55
Utilización de una biblioteca de modelos de tendencia para optimización	56
Mejora de la biblioteca de Modelos para Optimización	61
Errores de los Modelos para Optimización	62
2.4.2 Conclusiones	65
2.5 Bibliografía	68

Capítulo 3: OPTIMIZACION EXPERIMENTAL UTILIZANDO MODELOS DE TENDENCIA PROBABILISTICOS	
3.1 Hipótesis de Trabajo	71
3.2. Integración de bootstrapping en el desarrollo de modelos de tendencia probabilístico	71
3.3. Ejemplo 1: Diseño de Experimentos Dinámicos para la Optimización de la producción de penicilina en un cultivo de <i>Penicillium crysogenum</i> utilizando un modelo de tendencia probabilística. Modelo estructuralmente correcto	72
3.4. Diseño de Experimentos Dinámicos para la Optimización de la producción de penicilina en un cultivo de <i>Penicillium crysogenum</i> utilizando un modelo de tendencia probabilístico. Proceso estructuralmente diferente al modelo	80
3.5 Conclusiones	90
3.6 Bibliografía	91
Capítulo 4: CLONADO Y EXPRESIÓN DE rBm95 DE <i>Rhipicepalus microplus</i> EN <i>Escherichia coli</i>	
4.1. <i>Rhipicepalus microplus</i>	93
4.1.1. Problemática de la garrapata común del ganado	93
4.1.2. Control	95
4.1.3. Antígenos de <i>R. microplus</i> . Vacunas	95
4.2.1. Descripción físico-química de Bm95	96
4.3. Diseño de clonado de <i>bm95</i>	98
4.3.1. Diseño de la secuencia nucleotídica <i>bm95</i> para la expresión de rBm95 en <i>Escherichia coli</i>	98
4.3.2. Sistema de expresión utilizado. Breve descripción	99
4.3.3. Clonado y expresión de rBm95 en <i>E. coli</i>	102
4.3.4. Clonado de construcción geneCRISTALDI_pMK	103
4.3.5. Digestión y Fosforilación	104
4.3.6. Ligación	104
4.3.7. Clonado de construcción pET-24+-geneCRISTALDI	104
4.3.8. Secuenciación de la construcción pET-24+-geneCRISTALDI	105
4.3.9. Expresión de Bm95	106
4.3.10. Análisis de estabilidad de la construcción pET-24+- <i>bm95</i> en <i>E. coli</i> BL21(DE3)	109
4.4. Conclusiones	110
4.5. Bibliografía	111
Capítulo 5: OPTIMIZACIÓN EXPERIMENTAL DE PRODUCCIÓN DE RBM95 UTILIZANDO DISEÑO DE EXPERIMENTOS DINÁMICOS BASADO EN MODELOS DE TENDENCIAS PROBABILÍSTICOS	
5.1. <i>Escherichia coli</i> en cultivos de alta densidad celular.	114
5.1.1. Consideraciones generales.	114
5.1.2. <i>Scaling up</i>	118
5.2. Descripción del biorreactor	119
5.3. Definición de las condiciones de cultivo	121
5.3.1. Parámetros constantes de la política de operación.	121
5.3.2. Parámetros manipulables de la política de operación	123
5.4. Biblioteca de Modelos de Tendencias para Optimización	124
5.5. Mejora de la producción de rBm95	125
5.6 Conclusiones	138
5.7 Bibliografía	139
Capítulo 6: CONCLUSIONES Y PERSPECTIVAS FUTURAS	
6.1 Conclusiones generales	144
6.2 Perspectivas futuras	145
ANEXOS	
Anexo I: ANALISIS DE SENSIBILIDAD GLOBAL (GSA)	148
Error esperado debido a la fijación de parámetros en un modelo	150

Teorema 1	151
Análisis de Sensibilidad para modelos con parámetros correlacionados	152
Algoritmo computacional para el cálculo de los Índices de Sensibilidad Global	152
Bibliografía	154
Anexo II: BOOTSTRAPPING	156
Teoría	156
Teorema de Glivenko-Cantelli	158
Intervalos de confianza	159
Integración de <i>Bootstrapping</i> al Diseño de Experimentos Dinámicos basado en Modelos para Optimización	160
Bibliografía	161
Anexo III: Secuencias	162
Secuencia 1	163
Secuencia 2	164
Secuencia 3	165
Secuencia 4	166
Secuencia 5	167
Secuencia 6	168
Secuencia 7	169
Secuencia 8	170
Anexo IV: Técnicas microbiológicas y de Biología Molecular	
A.4.1 Medios de cultivo	189
A.4.2 Transformación de ADN en <i>E. coli</i>	189
Preparación de células de <i>E. coli</i> competentes	189
Transformación de células bacterianas competentes	189
Selección de los clones transformantes de <i>E. coli</i>	190
A.4.3 Miniprep de ADN plasmídico	190
A.4.4 Digestión del ADN con Endonucleasas de Restricción.	190
A.4.5 Electroforesis de ácidos nucleicos en geles de agarosa.	190
A.4.6 Fosforilación T4 kinasa	191
A.4.7 Electroforesis de proteínas en geles de poliacrilamida (SDS- PAGE).	191
Tinción de proteínas en los geles de poliacrilamida.	192
A.4.8 Purificación de rBm95 con perlas de agarosa con quelación de Ni ²⁺	192
A.4.9 Evaluación de estabilidad de construcción plasmídica pET-24b- <i>bm95</i> en cepas de <i>E. coli</i> transformadas.	192
Procedimiento de Recuento en placa en superficie	193
Anexo V: Técnicas Operatorias para la determinación de variables de estado del cultivo en biorreactor BioFlo 110.	
A.5.1 Determinación de concentración celular	194
A.5.2 Determinación de glucosa por kit enzimático	194
A.5.3 Determinación de proteínas por el método de Lowry	195
A.5.4 Determinación de acetato por kit enzimático	196

Índice de Tablas

Tabla	Página
2.3.1 Intervalos de incertidumbre del modelo de Menezes (Menezes y col., 1994) y parametrización del biorreactor <i>in silico</i>	38
2.3.2 Parámetros de la política de operación para caudal de alimentación constante	39
2.3.3 Índices de sensibilidad para los parámetros del modelo adoptando una alimentación constante	40
2.3.4 Valores reales y re-estimados de los parámetros más relevantes según GSA	40
2.3.5 Parámetros de la política de operación para caudal de alimentación siguiendo un polinomio lineal inverso	41
2.3.6 Índices de sensibilidad normalizados para los parámetros del modelo adoptando una alimentación siguiendo un polinomio lineal inverso	42
2.3.7 Valores reales y re-estimados de los parámetros más relevantes según GSA	42
2.3.8 Parámetros de la política de operación para caudal de alimentación siguiendo un polinomio inverso cuadrático	45
2.3.9 Índices de sensibilidad normalizados para los parámetros del modelo adoptando una alimentación siguiendo un polinomio inverso cuadrático.	46
2.3.10 Valores reales y re-estimados de los parámetros más relevantes según GSA	46
2.3.11 Parámetros de la política de operación para caudal de alimentación siguiendo un polinomio inverso de 3° grado.	50
2.3.12 Índices de sensibilidad normalizados para los parámetros del modelo adoptando una alimentación siguiendo un polinomio inverso de 3° grado	50
2.3.13 Valores reales y re-estimados de los parámetros más relevantes según GSA	51
2.4.1 Intervalos de incertidumbre del modelo de Menezes (Menezes, 1994) y parametrización del biorreactor <i>in silico</i>	58
2.4.2 Parámetros de la política de operación para caudal de alimentación constante	58
2.4.3 Perfil de los índices de sensibilidad normalizados de los parámetros durante la optimización (iter. 1-3)	64
2.4.4 Perfil de los índices de sensibilidad normalizados de los parámetros durante la optimización (iter. 4 y 5)	64
2.4.5 Valores re-estimados de los parámetros más relevantes según GSA	65
2.4.6 Errores paramétrico, estructural y total para la biblioteca de modelos (iter. 1-3)	67
2.4.7 Errores paramétrico, estructural y total para la biblioteca de modelos (iter. 4 y 5)	67
3.3.1 Parámetros de la política de operación durante optimización experimental y los de la política de operación óptima	74
3.3.2 Índices de sensibilidad para los parámetros del modelo	75
3.3.3 Rangos iniciales y re-estimados de los parámetros del modelo durante la optimización	76
3.4.1 Parametrización del biorreactor <i>in silico</i>	81
3.4.2 Parámetros de la política de operación durante optimización experimental y los de la política de operación óptima	82
3.4.3 Índices de sensibilidad para los parámetros del modelo	84
3.4.4 Rangos iniciales y re-estimados de los parámetros del modelo durante la optimización	87
4.1 Cantidad y abundancia relativa de residuos aminoácidos en Bm95	97
4.2 Secuencia y parámetros físico-químicos del cebador diseñado	105
4.3 Estabilidad de la construcción pET-24+ <i>-bm95</i> en <i>E.coli</i> BL21(DE3)	110
5.1 Concentraciones máximas pre-inhedorias de nutrientes en cultivos de <i>E. coli</i>	116
5.2 Sistemas de control del biorreactor BioFlo 110	121
5.3 Parámetros de la política de operación fijados <i>a priori</i> de la optimización	123
5.4 Valores límites de los parámetros manipulables de la política de operación	124
5.5 Políticas de operación calculadas (cursivas) e implementadas durante la optimización	126
5.6 Errores paramétricos, estructurales y totales	130
5.7 Índices de sensibilidad global de primer orden (experiencias 0 y 1)	130
5.8 Índices de sensibilidad global de primer orden (experiencias 2 y 3)	130
5.9 Intervalos de confianza iniciales (95 percentilo) y valor más frecuente de los parámetros de cada modelo de la biblioteca de modelos de tendencias	131

5.10	Intervalos de confianza (95 percentilo) y valor más frecuente de los parámetros de cada modelo de la biblioteca de modelos de tendencias re-estimados en las experiencias 1 y 2	134
5.11	Intervalos de confianza (95 percentilo) y valor más frecuente de los parámetros de cada modelo de la biblioteca de modelos de tendencias re-estimados en la experiencia 3	134
A.4.1.1	Medio LB selectivo (Luria Bertani)	182
A.4.1.2	Medio TB selectivo (Terrific Broth)	182
A.4.4.1	Mezcla de reacción digestión	184
A.4.7.1	Marcador de Peso Molecular	185
A.5.2.1	Procedimiento para determinación de glucosa	187
A.5.3.1	Procedimiento para determinación de proteínas totales	189
A.5.4.1	Procedimiento para la determinación de acetato	189

Índice de Figuras

Figura	Página
1.1	10
1.2	12
2.1	18
2.2	20
2.3a	26
2.3b	32
2.3.1	43
2.3.2	43
2.3.3	44
2.3.4	44
2.3.5	47
2.3.6	48
2.3.7	48
2.3.8	48
2.3.9	51
2.3.10	52
2.3.11	52
2.3.12	52
2.4.1	59
2.4.2	60
2.4.3	60
2.4.4	60
3.1	72
3.3.1	75
3.3.2	77
3.3.3	79
3.3.4	79
3.4.1	83
3.4.2	86
3.4.3	86
3.4.4	88
3.4.5	88

3.4.6	<i>fdp</i> empírica de $\zeta_{m\acute{a}x}$ (a) inicial y (b) iteración 2	89
4.1	<i>R. microplus</i> hembra (vista ventral)	93
4.2	Posibles zonas de desarrollo de <i>R. microplus</i>	93
4.3	Cuello de bovino con infestación aguda de <i>R. microplus</i>	94
4.4	Representación del vector de expresión pET-24 http://www.merckbiosciences.com/docs/NDIS/TB072-000.pdf	101
4.5	Funcionamiento general del sistema DE3-pET. (pET System Manual, Novagen)	102
4.6	Mapa del plásmido proporcionado por la empresa GenArt SA	103
4.7	Gel de agarosa 1% de minipreps de construcción GenArt (calle 4) y pE T-24b+ (calle 5).	104
4.8	Gel de agarosa 1% de minipreps de colonias de <i>E. coli</i> Top10 transformantes digeridas: MW ladder (calle 1), colonia negativa (calle 2), colonia positiva (calle 3), fragmento geneCRISTALDI aislado (calle 4), pET-24+ digerido con NdeI y HindIII (calle 5).	105
4.9	SDS-PAGE revelado con Coomassie Brilliant Blue de cultivos de <i>E. coli</i> BL21 transformada con la construcción pET-24b- <i>bm95</i> : calle 1, 2 y 3: cultivo en caldo LB sin inducir, luego de 3 horas de inducción, inducido toda la noche, respectivamente; calle 4: MW ladder; calle 5, 6 y 7: cultivo en caldo TB sin inducir, luego de 3 horas de inducción, inducido toda la noche, respectivamente.	107
4.10	SDS-PAGE revelado con Coomassie Brilliant Blue (calles 1-5) y Western Blot con anti-Histag irradiado con luz UV de cultivos de <i>E. coli</i> BL21 transformada con la construcción pET-24b- <i>bm95</i> : calles 1 y 6: sin inducir; calles 2-4 y 7-10: inducidos.	107
4.11	SDS-PAGE revelado con Coomassie Brilliant Blue de cultivos de <i>E. coli</i> BL21 transformada con la construcción pET-24b- <i>bm95</i> : calle 1: MW ladder; calle 2-5: fracción insoluble de cultivo sin inducir y 1, 2 y 3 hs post inducción, respectivamente; calle 6-9: fracción soluble de cultivo sin inducir y 1, 2 y 3 hs post inducción, respectivamente.	108
4.12	SDS-PAGE revelado con Coomassie Brilliant Blue de cultivos de <i>E. coli</i> BL21 transformada con la construcción pET-24b- <i>bm95</i> : calle 1: MW ladder; fracción insoluble de cultivo sin inducir: sobrenadante (calle 2) y elusión (calle 5); fracción insoluble de cultivo inducido (4h): sobrenadante (calle 3) y elusión (calle 6); fracción soluble de cultivo inducido (4h) : sobrenadante (calle 3) y elusión (calle 6).	109
5.1	Ilustración del sistema BioFlo 110	119
5.2	Vista del sistema BioFlo 110 montado en laboratorio	120
5.3	Dinámica de oxígeno para cultivo representativo de <i>E. coli</i> BL21 transformada con la construcción pET-24b- <i>bm95</i>	122
5.4	Cambio metabólico en <i>E. coli</i> BL21 transformada con la construcción pET-24b- <i>bm95</i> debido a la inducción	123
5.5	Políticas de alimentación del biorreactor BioFlo 110 durante el proceso de optimización	127
5.6	Comparación de Políticas de alimentación diseñadas en la experiencia 2 y la considerada de convergencia	128
5.7	Trayectorias de la función objetivo <i>J</i> y las estimaciones de los modelos Monod (verde), Contois (rojo), Mantenimiento (celestes), 2° Orden (violeta) para las experiencias: 0, 1, 2, 3.	131
5.8	<i>fdp</i> empíricas iniciales de los parámetros del modelo Contois	132
5.9	<i>fdp</i> empíricas iniciales de los parámetros del modelo 2° orden	133
5.10	<i>fdp</i> empíricas para el parámetro k_{lis} del modelo Contois estimadas luego de las experiencias inicial (a), 1 (b) y 2 (c)	135
5.11	<i>fdp</i> empíricas para el parámetro k_{lis} del modelo 2° orden estimadas luego de las experiencias inicial (a), 1 (b) y 2 (c)	135
5.12	Evolución de espacio de estimaciones de <i>J</i> del modelo Contois durante la optimización	136
5.13	Evolución de espacio de estimaciones de <i>J</i> del modelo 2° orden durante la optimización	137

Resumen

En esta tesis se aborda la problemática de la mejora de la *performance* de bioprocesos y se desarrolla una metodología para la optimización experimental, la cual es finalmente implementada para la mejora de la producción del antígeno rBm95 en un cultivo de alta densidad celular de *Escherichia coli* a escala banco.

La metodología plantea la búsqueda de las condiciones de operación óptimas mediante un ciclo iterativo que involucra el Diseño de Experimentos Dinámicos para Optimización. Los experimentos para optimización son realizados bajo las condiciones de operación que se suponen harán máximo el índice de *performance* del proceso, junto con una estrategia de muestreo que permitirá la obtención de datos más informativos. Los experimentos son calculados utilizando modelos de tendencias que describen la dinámica general del bioproceso y que son mejorados en cada ciclo mediante la estimación adecuada de sus parámetros cinéticos más relevantes en la estimación de la trayectoria de la función objetivo.

Durante el proceso de optimización, la incertidumbre asociada al uso de modelos de tendencias es analizada mediante la discriminación entre las fuentes de incertidumbre estructural y paramétrica, y para su cuantificación son propuestas formulaciones sencillas. Además, son discutidas las condiciones que deben cumplirse para la convergencia al óptimo de la metodología.

Para mostrar las propiedades y comportamiento de la metodología propuesta, el Diseño de Experimentos Dinámicos basado en Modelos para Optimización es aplicado primero a diversos ejemplos de bioprocesos *in silico* con diferentes grados de dificultad, y luego a la optimización experimental de la producción del antígeno rBm95 que abrió las puertas para el control de *R. microplus* en nuestro país.

R. microplus es considerada la garrapata parasitaria más importante en el mundo ya que ataca a ganados tan diversos como bovinos, ovinos, búfalos, burros, entre otros. Este insecto afecta la producción de carne y leche, las tasas de preñez, puede ocasionar cuadros de anemia más o menos severos, o incluso la muerte del animal parasitado. Por otro lado, las heridas derivadas de las picaduras en la piel son potenciales centros de miasis (bicheras) y disminuyen la calidad de los cueros. Además, genera daños ocultos ya que *R. microplus* es el vector de babesiosis (causada por los protozoos *Babesia bigemina* y *B. bovis*) y anaplasmosis (causada por *Anaplasma marginale*).

Ha sido obtenido un clon de *E. coli* capaz de expresar satisfactoriamente el antígeno recombinante Bm95 de *Rhipicepalus microplus*, el cual ha demostrado ser capaz de generar una protección inmune contra este ácaro gigante cuando se lo utiliza en formulaciones vacunales para el ganado. La metodología de optimización desarrollada es finalmente implementada con una biblioteca de cuatro modelos de tendencias para la optimización de la producción de rBm95 de un cultivo de alta densidad celular de *E. coli* en un biorreactor a escala banco.

Introducción

Las empresas de los sectores farmoquímico y biotecnológico afrontan singulares desafíos y oportunidades para integrar propuestas innovadoras a sus carteras de productos y procesos de elaboración asociados. El alto valor de los productos amplía significativamente el universo de metodologías químicas y biotecnológicas que *a priori* pueden ser consideradas económicamente viables (Schultz y Krska, 2007). Por esta razón estas industrias son adoptantes tempranos de métodos innovadores, generalmente en fase embrionaria. Sin embargo, en recientes años, una creciente competencia está ejerciendo una gran presión colocándolas en la necesidad de penetrar de manera rápida el mercado objetivo. Esto significa menos tiempo para explorar y desarrollar procesos escalables y robustos (Pisano, 1997). Esta presión es principalmente sufrida porque los procesos de manufactura son mantenidos en estado embrionario hasta etapas tardías del desarrollo debido a la alta tasa de fracaso de drogas candidatas que caracteriza los estadios avanzados de su desarrollo y las fases de pruebas clínicas. Inmediatamente después, la droga aprobada tiene que estar comercialmente disponible para tomar ventajas durante el limitado período de protección otorgado por la Patente de Invención (Minnaard y col., 2007). Esto deja un margen muy estrecho para el desarrollo de los procesos industriales y, en consecuencia, exige el escalado rápido y además económico de los procesos involucrados (Walsh, 2007).

En ese escenario dinámico, caracterizado por el riesgo técnico y económico inherente a la innovación, el posicionamiento estratégico de las empresas depende de la capacidad de anticipación tecnológica y del lanzamiento comercial temprano. Como regla, las fallas técnicas son una posibilidad cierta y el éxito comercial se encuentra lejos de estar garantizado aún cuando aquellas estén resueltas. De ahí que el diseño de productos y procesos de producción debieran estar sincronizados en una estrategia interactiva que utilice en forma eficiente la información que suele ser escasa y con grandes niveles de incertidumbre (Bonvin, 1998; Visser y col., 2000; Martínez y Wilson, 2003). Un conocimiento imperfecto o incompleto de los fenómenos fundamentales involucrados y la presencia de múltiples factores que no pueden ser controlados o compensados es un problema central en el cambio de escala no sólo por los aspectos económicos sino también por los riesgos de excursiones técnicas (Stephanopoulos y col., 1999), de afectar propiedades de uso final del producto obtenido o afectar negativamente las operaciones downstream (Terwiesch y col., 1994). En consecuencia, las políticas de

operación utilizadas a escala industrial son en extremo conservadoras y los rendimientos obtenidos son muy inferiores a los que podría proporcionar una política de operación optimizada.

De lo anterior deviene la importancia de contar con herramientas de optimización que asistan el desarrollo, cambio de escala y/o mejora del proceso de producción colocándolo en el límite superior de desempeño, en corto período y forma confiable, usando mínima información.

La filosofía subyacente al presente trabajo de tesis es la de aceptar que los modelos utilizados para describir el comportamiento dinámico del sistema estudiado son simplificaciones de la realidad pero que aún así pueden asistir en la mejora de la *performance* de un proceso innovador. Esta idea utilitarista hace un uso exhaustivo de las herramientas disponibles, los Modelos de Tendencias, para lograr la mejor política de operación posible, eludiendo el costoso y prolongado proceso de formulación de modelos fundamentales.

Los objetivos de carácter metodológico de esta tesis son:

- Implementar nuevas metodologías de Análisis de Sensibilidad Global (GSA) para la exploración de la sensibilidad de los modelos matemáticos a la incertidumbre de sus parámetros para seleccionar los más relevantes, en cuya mejor estimación se debería invertir tiempo y dinero.
- Implementar computacionalmente un algoritmo completo para el Diseño de Experimentos Dinámicos basados en Modelos de Tendencias para Optimización que cuente con una biblioteca de modelos alternativos a partir de los cuales pueda ser escogido el más adecuado para realizar el cálculo de los experimentos para optimización.
- Realizar un estudio exhaustivo de las incertidumbres estructural y paramétrica de los modelos para optimización y proponer formulaciones para su cuantificación y caracterización mediante la utilización de *bootstrap*.

En su desarrollo, las distintas fases metodológicas del trabajo son intencionalmente aplicadas a sistemas que guardan una relación con un creciente nivel de desafío y con una trayectoria que aborda el modelado cinético de procesos químicos que involucran sistemas de reacciones relativamente complejos, la optimización de bioprocesos *in silico*, y finalmente la optimización de la producción en cultivo de alta densidad celular

de *E. coli*, en un biorreactor a escala banco, del antígeno rBm95 contra *R. microplus* considerada la garrapata parasitaria más importante en el mundo. Como fue mencionado anteriormente, este insecto afecta la producción de carne y leche, a las tasas de preñez, ocasiona cuadros de anemia más o menos severos o incluso la muerte del animal parasitado. Por otro lado, las heridas derivadas de las picaduras en la piel son potenciales centros de miasis (bicheras) y disminuyen la calidad de los cueros. Además, genera daños ocultos ya que *R. microplus* es el vector de babesiosis (causada por los protozoos *Babesia bigemina* y *B. bovis*) y anaplasmosis (causada por *Anaplasma marginale*). De ahí deviene la importancia de contar con una vacuna eficaz para hacer frente a infestaciones recurrentes sobre todo teniendo en cuenta que la especie argentina de ese parásito presenta resistencia a las vacunas que actualmente pueden encontrarse en el mercado (Gavac® y TickGuard®). Los precios de mercado de un producto con esas características no suelen dejar grandes márgenes de ganancia y por lo tanto aparece la necesidad de contar con una metodología capaz de incrementar en forma rápida la productividad de un biorreactor para la producción de rBm95.

Los objetivos de interés tecnológico de esta tesis son:

- Obtener un clon de *E. coli* capaz de expresar satisfactoriamente el antígeno recombinante Bm95, específico de la cepa autóctona de *R. microplus*, el cual ha demostrado ser capaz de generar una protección inmune contra este ácaro gigante cuando se lo utiliza en formulaciones vacunales para el ganado.
- Desarrollar una metodología experimental e iterativa de optimización basada en modelos de tendencias y aplicarla exitosamente en la mejora de la *performance* de un biorreactor de alta densidad celular para la producción de rBm95.

Desde esta motivación, es de particular interés:

- Obtener la secuencia de la proteína Bm95, realizar un estudio genómico de compatibilidad de codones respecto al sistema de expresión a utilizar y de optimización de la secuencia.
- Sintetizar el gen *bm95* con sitios de restricción adecuados para su clonado en un vector de expresión.
- Insertar el gen *bm95* en el vector de expresión pET-24b, clonarlo y expresar rBm95 en diferentes cepas de *E. coli*.

- Desarrollar y poner a punto técnicas de cuantificación de las variables *on-line* y *off-line* que serán utilizadas en cada iteración para disminuir la incertidumbre sobre la localización del óptimo.
- Optimizar la producción del antígeno rBm95 en cultivo de alta densidad celular de *E. coli* utilizando Diseño de Experimentos Dinámicos basados en Modelos para Optimización en biorreactor a escala banco.

Brevemente, en cada iteración del procedimiento propuesto para la optimización de la producción de rBm95, el objetivo es seleccionar el mejor modelo de una biblioteca de modelos de tendencias para utilizarlo en el diseño de un nuevo experimento dinámico que mejore la *performance*. Una vez que el nuevo experimento para optimización ha sido diseñado, la política de operación es implementada y nuevos datos son obtenidos durante el ensayo experimental siguiendo la estrategia de muestreo óptima. Dichos datos son utilizados para reducir la incertidumbre paramétrica en un esquema bayesiano de reducción del espacio paramétrico de los modelos de tendencias que componen la biblioteca. Esto es logrado vía la reestimación de un subconjunto de parámetros seleccionados por la incidencia de su desconocimiento sobre la variabilidad de las predicciones de cada modelo mediante un Análisis de Sensibilidad Global (GSA). Una vez que los modelos han sido actualizados/mejorados, los errores asociados a la incertidumbre estructural y paramétrica son calculados para elegir al mejor modelo para optimización y así comenzar una nueva iteración.

Respecto de la organización del trabajo de tesis se puede destacar, a modo de guía para el lector, vale mencionar lo siguiente:

- En el capítulo I son explicitados conceptos generales relacionados con el modelado matemático y los análisis de sensibilidad de los modelos con el propósito de uniformar los conceptos que posteriormente serán utilizados en el trabajo. A modo de ejemplo y como muestra de la utilidad de la herramienta, el GSA es integrado en un esquema para la formulación del modelo cinético más sencillo que describe la hidrogenación catalítica de Avermectina para la preparación de Ivermectina. El modelo sencillo seleccionado describe satisfactoriamente el proceso llevado a cabo en un reactor industrial, y su búsqueda fue motivada por la utilidad para integrarlo a métodos de simulación y control del proceso industrial.

- En el capítulo II es formalmente presentada la Metodología del Diseño de Experimentos Dinámicos para Optimización basada en Modelos de Tendencias y son discutidas las condiciones de convergencia, estabilidad del método, y otros conceptos relacionados. Además, se propone el desacople y cuantificación de las incertidumbres estructural y paramétrica de los modelos utilizando medidas de dichos errores basadas en GSA. Dos ejemplos de optimización de bioprocesos *in silico* son presentados. En uno de ellos son analizados los resultados en función de la manera que es definida la política de operación de un bioproceso. En el otro se ejemplifica la optimización de un biorreactor utilizando una biblioteca de modelos de tendencias.
- En el capítulo III se presenta una mejora del método de optimización mediante la incorporación de la técnica de *bootstrap* al ciclo iterativo para una mejor caracterización y conocimiento del dominio paramétrico de los modelos de tendencias. Este aporte es analizado mediante ejemplos de optimización de biorreactores *in silico* en los que la incertidumbre estructural se encuentra ausente o presente, respectivamente.
- El capítulo IV está dedicado a la presentación del problema tecnológico asociado a la producción del antígeno rBm95. Además, es realizada una caracterización molecular y estructural de dicha proteína, a la vez que se describe la modificación y obtención de la cepa *E. coli* capaz de expresarlo satisfactoriamente.
- En el capítulo V la implementación práctica del Diseño de Experimentos Dinámicos para Optimización basado en Modelos de Tendencias es ejemplificada mediante su aplicación exitosa a la mejora de la *performance* de un biorreactor de alta densidad celular para la producción de rBm95, a escala banco.

Finalmente, se explicitan las conclusiones generales a las que se han arribado en el marco de la realización de esta tesis y se sugieren posibles avances futuros en la temática.

Con la intención de facilitar la lectura, los capítulos en los que aparecen tanto técnicas numéricas como experimentales remiten a Anexos específicos en los que se ha sistematizado información relevante y detalles.

Bibliografía.

Bonvin, D. (1998). Optimal operation of batch reactors: a personal view. *Journal of Process Control*, **8**(5,6), 355-368.

Martínez, E. C.; Wilson, J. A. (2003). Evolutionary optimization of batch process systems using imperfect models. *Proceeding of the Indian Natural Science Academy*, **69**, 403-428.

Minnaard A.; Feringa B.; Lefort L.; Vries J. (2007). Asymmetric Hydrogenation Using Monodentate Phosphoramidite Ligands. *Accounts of Chemical Research*, **40**(12), 1267–1277.

Pisano G. (1997). *The development factory. Unlocking the potential of process innovation*. Howard Bussines School Press, Boston, Massachussts.

Schultz C.; Krska S. (2007). Unlocking the potential of asymmetric hydrogenation at Merck. *Accounts of Chemical Research*, **40**(12), 1320–1326.

Terwiesch P.; Agarwal M.; Rippin D. (1994). Batch unit optimization with imperfect modelling: a survey. *Journal of Process Control*, **4**(4), 238-258.

Visser, D.; van der Heijden, R.; Mauch, K.; Reuss, M.; Heijnen, S. (2000). Tendency modeling: A new approach to obtain simplified kinetic models of metabolism applied to *Saccharomyces cerevisiae*. *Metabolic Engineering*, **2**(3), 252-275.

Walsh, G. (2007). *Pharmaceutical biotechnology: concepts and applications*; John Wiley & Sons Ltd, Chichester, England.

“Hablando por medio de paradojas: el hombre es un ser limitado que no tiene límites. Traspasar los límites es algo innato al hombre. Para traspasar los límites es necesario tenerlos, es necesario tener lo real y lo posible, lo determinado y lo indeterminado. El mundo de cada animal es completo y seguro, nosotros sabemos que estamos limitados, y para eso tenemos que traspasar los límites, haberlos visto desde afuera”

Simmel

1.1. Modelos

1.1.1. Concepto

La real academia española acepta diferentes acepciones para la palabra “modelo”. Las más comúnmente utilizadas son:

1. m. Arquetipo o punto de referencia para imitarlo o reproducirlo.
2. m. En las obras de ingenio y en las acciones morales, ejemplar que por su perfección se debe seguir e imitar.
3. m. Representación en pequeño de alguna cosa.
4. m. Esquema teórico, generalmente en forma matemática, de un sistema o de una realidad compleja, (...) que se elabora para facilitar su comprensión y el estudio de su comportamiento.
5. m. Objeto, aparato, construcción, etc., o conjunto de ellos realizados con arreglo a un mismo diseño.
6. m. Vestido con características únicas, creado por determinado modista, y, en general, cualquier prenda de vestir que esté de moda.
7. m. En empresas, u en aposición para indicar que lo designado por el nombre anterior ha sido creado como ejemplar o se considera que puede serlo.¹

En esta tesis, la palabra “modelo” es utilizada bajo la acepción 4; aceptando que los modelos matemáticos son simplificaciones de una realidad compleja y, casi por definición inefable, que ayudan a la comprensión de los fenómenos que suceden en el sistema estudiado y mediante los cuales intentamos predecir su comportamiento.

1.1.2. Estructura y parámetros. Incertidumbre estructural y paramétrica

La forma en que se cree que la realidad funciona, o como se creía “las leyes universales que rigen el comportamiento de los cuerpos” (sic), codificada matemáticamente como una función $f(\bullet)$ constituye la estructura del modelo. Dicha función relaciona las variables que definen el estado del sistema mediante factores de escala o normalización (asumidos constantes) llamados parámetros.

¹ http://buscon.rae.es/draeI/SrvltConsulta?TIPO_BUS=3&LEMA=cultura

La praxis nos muestra que esa pretendida “universalidad” buscada por las personas que construyen modelos es, al menos, una pretensión ingenua ya que existen ejemplos en los cuales las mismas observaciones pueden ser “explicadas” o descritas mediante ecuaciones o sistemas de representación irreconciliables entre sí. Por lo tanto, la formulación matemática del modelo constituye una fuente de incertidumbre *per se* y se conoce como **incertidumbre estructural**.

Como lo han expresado Saltelli y col. (Saltelli y col., 2008) “desde la época de Galileo, los científicos han tenido que lidiar con la capacidad limitada de la mente humana para crear mapas útiles del ‘mundo’ en forma de ‘modelos’. La emergencia de ‘leyes’ puede ser vista en este contexto como el resultado de un fatídico proceso de simplificación, separación e identificación que lleva a un modelo de simpleza y belleza poco comunes”².

Es por eso que hoy en día la mayoría de los científicos que construyen modelos estarían de acuerdo en afirmar que los modelos no pueden ser “validados”, en el sentido de probados como ciertos, sino más bien que pueden ser “corroborados” ya sea mediante ensayos formales de consistencia o por su capacidad para explicar o predecir nuestro “mundo”³. Como esa corroboración es concomitantemente parcial, los modelos son clasificados por su valor heurístico: “los modelos pueden corroborar una hipótesis...los modelos pueden resolver discrepancias con otros modelos. Los modelos pueden ser usados para Análisis de Sensibilidad – para responder a preguntas ‘que tal si?’ –”⁴ (Oreskes, 2004), pero no son susceptibles de ser probados.

Lo que hace al modelado y a la investigación científica tan ominoso es la incertidumbre. Esto no es un accidente del método científico sino su esencia, ya que en la ciencia tanto las preguntas como las respuestas están vinculadas por medio de ella.

Los modelos matemáticos que serán encontrados en el presente trabajo intentan describir procesos químicos y/o bioquímicos y tienen la forma general:

$$\frac{dx}{dt} = f(x(t), \varphi(w, t), \theta, t) \quad 0 \leq t \leq t_f, \quad x(0): \text{dado} \quad (1.1)$$

donde $x(t)$ es un vector n_s -dimensional de variables de estado dependientes del tiempo, $w \in W$ es un vector m -dimensional de parámetros que definen la política de operación

² Saltelli y col.; Global Sensitivity Analysis. The primer.; 2008

³ Idem 2.

⁴ Oreskes; Risk Analysis; 2004.

del proceso considerado representada por $\wp(w, t)$ y W es el espacio de valores que pueden adoptar esos parámetros ya sea porque existen restricciones físicas y/o operativas (temperaturas máximas y/o mínimas, volumen máximo de trabajo, caudal máximo de una bomba, etc.) o económicas (carga máxima de algún reactivo costoso, tiempo total del proceso, etc.), $\theta \in \Theta$ es un vector p -dimensional de parámetros del modelo que están definidos en el espacio de incertidumbre paramétrico Θ , y t_f es el tiempo final durante el cual el comportamiento de las variables que describen el proceso intenta ser estimado.

Cuando una estructura de estos modelos es adoptada, Θ representa la incertidumbre de las estimaciones atribuida al desconocimiento de los valores verdaderos de los parámetros, a la que denominaremos **incertidumbre paramétrica**. Por lo tanto asumiendo una estructura correcta del modelo, y por ende en ausencia de incertidumbre estructural, el problema de predicción se reduce a encontrar los valores verdaderos de θ mediante la reducción de la incertidumbre paramétrica.

1.1.3. Clasificación

Existen múltiples formas de clasificar a los modelos dependiendo de qué criterio se tome para realizar la diferenciación. Una manera puede ser la que los divide en *law-driven* y *data-driven* (Georgakis, 1995; Martínez y col, 2009). Los primeros son formulados a partir de leyes que se supone el sistema debe seguir, mientras que los otros intentan describir estadísticamente el comportamiento de una señal medida con cierta frecuencia durante la evolución del sistema estudiado.

Los partidarios de los modelos *data-driven* argumentan que éstos pueden ser contruidos de manera parsimoniosa, es decir que pueden describir la realidad con un mínimo de parámetros ajustables. En cambio, los modelos *law-driven* suelen estar sobrep parametrizados, en el sentido que pueden contener más leyes de las que los datos recolectados podrían confirmar, y por lo tanto, presentar problemas de identificabilidad y observabilidad (Asprey y Macchietto, 2000). Sin embargo, es esa característica mecanicista la que potencialmente les conferiría la capacidad de describir el sistema en condiciones no observadas experimentalmente (extrapolación) ya que los modelos *data-driven* son confiables sólo bajo las condiciones en las cuales los datos utilizados para su construcción han sido adquiridos (interpolación) y, consiguientemente, la posibilidad de utilizarlos para realizar extrapolaciones sería reducida.

1.2. Análisis de Sensibilidad de Modelos

1.2.1. Generalidades

Un análisis de sensibilidad (AS) de un modelo dado podría ser definido como un conjunto de métodos, que pueden ser numéricos o analíticos, con los que se intenta conocer y, más precisamente, cuantificar cómo una fuente de incertidumbre en los *inputs* produce una variación en los *outputs* del modelo (Saltelli y col., 2004; 2006; Stirling, 2000). Esta clase de análisis es útil porque puede revelar errores en la estructura del modelo, identificar regiones críticas en el espacio de los *inputs* con el objetivo de establecer prioridades para la investigación y realizar simplificaciones, entre otras.

Un AS es considerado por muchos expertos como un prerrequisito a la hora de construir un modelo, cualesquiera sea su finalidad y/o campo de aplicación. Este análisis debe estar integrado en la actividad misma del modelado ya que se apela a los AS cuantitativos a la hora de defender la calidad de las conclusiones extraídas a partir de un modelo. Tanto es así que el Departamento de Trabajos Públicos de Holanda ha incluido este tipo de herramientas en su Good Modelling Practice Handbook (GMoP) (STOWA/RIZA, 1999).

Hace varios años, Hornberger y Spear aseguraron que: “la mayoría de los modelos de simulación serán complejos, con muchos parámetros, variables de estado y relaciones no lineales. En las mejores condiciones esos modelos tienen muchos grados de libertad y, con juguetes ‘juiciosos’, pueden producir virtualmente cualquier comportamiento deseado, frecuentemente tanto con estructuras como con parámetros plausibles”⁵. Este problema fue el que llevó al economista Leamer en 1990 a proponer un AS que llamó “Análisis de Sensibilidad Global” en el que supuestos alternativos son seleccionados y el correspondiente intervalo de inferencias es identificado. Las conclusiones son juzgadas como robustas sólo si el espacio de los supuestos es suficientemente amplio como para ser creíble y el correspondiente intervalo de inferencias es suficientemente estrecho como para ser útil”⁶.

El problema de la incertidumbre tanto en los *inputs* del modelo como en sus estimaciones es conocido como GIGO (*garbage in – garbage out*), y hasta existe un “principio” que dice: “La precisión en la salida aumenta a medida que la exactitud en

⁵ Saltelli, Risk Análisis, 2002.

⁶ Leamer, 1990.

las entradas disminuye”⁷, aludiendo a que la precisión en las *outputs* es mejorada sólo después de restringir arbitrariamente el espacio de las *inputs*.

1.2.2. Propiedades de un AS ideal

Un AS ideal debería tener las siguientes características generales:

- a. La capacidad de manejar la influencia de la escala y forma de los *inputs*. Esto significa que pueda extraer conclusiones válidas independientemente del rango de variación de los *inputs* y sus respectivas funciones de densidad de probabilidad (*pdf*, siglas en inglés).
- b. Incluir el fenómeno del promedio multidimensional. Esta característica es propia de los métodos globales ya que ellos evalúan el efecto de un *input* cuando a todos los demás también se les permite variar.
- c. Ser independiente del modelo, de tal forma que sea capaz de dilucidar las interacciones entre *inputs* (tan frecuentes en modelos no aditivos y no lineales).
- d. La capacidad de posibilitar la evaluación de un grupo de *inputs* como si fuese un único *input* (esto permite agilizar la interpretación de resultados).

1.2.3. Implementación de AS

Un algoritmo general para la implementación de un AS, independientemente tanto del modelo considerado como del tipo de AS ha realizar, contemplaría los siguientes pasos (Sainana, 2005; Borgonovo, 2003):

- a. Establecer unívocamente el objetivo del AS a través de la definición adecuada de una función *output* que dé respuestas a las preguntas que necesitamos contestar, esto es definir “el propósito” del modelo.
- b. Decidir cuáles son los *inputs* que serán incluidos en el AS.
- c. Definir las distribuciones de densidad de probabilidad de los *inputs* ya sea a partir de datos bibliográficos, de la construcción de distribuciones empíricas a partir de datos experimentales, de la consulta a expertos, definiendo una estructura de la correlación entre los *inputs*.
- d. Elegir un AS adecuado teniendo en cuenta:
 - Las preguntas que se desean responder.
 - La cantidad de simulaciones que pueden ser realizadas en base al costo computacional de cada una de ellas.

⁷ Stirling, 2000.

- La presencia de correlación entre las *inputs* magnitud de la cual podría reducir la complejidad del análisis.
- e. Generar la muestra de los *inputs* que estará formada por conjuntos de valores de los *inputs* con los que el modelo será evaluado.
 - f. Realizar las simulaciones a partir de la muestra de manera de tener N valores del *output* definido según el ítem a.
 - g. Analizar *outputs* y obtener las conclusiones buscadas o aquellas que llevarían a revisar alguno(s) de los pasos anteriores para dar inicio a una nueva iteración de AS.

Un inconveniente comúnmente encontrado en los AS es definir qué significa que un *input* sea relevante o importante en su efecto sobre el *output* del modelo. Ese concepto debe ser definido en el contexto de cada análisis ya que generalmente la repercusión de los *inputs* considerados individualmente tiene una distribución muy asimétrica respecto al *output* siendo que unos pocos logran explicar la mayor parte de la incertidumbre de las estimaciones y la mayoría tiene un efecto despreciable.

1.2.4. AS local y global

Supongamos que $Y = f(X_i), i = 1, 2, \dots, k$ es la representación matemática o computacional del modelo que desea ser estudiado, donde X_i es el i -ésimo *input* cuyo valor exacto es desconocido pero que, sin embargo, se cuenta con cierta información sobre su rango de variación y su *pdf* ya sea a través de su medición directa, la opinión de expertos o por su analogía con *inputs* de problemas similares. Toda esa información define el espacio Θ Y es el *output* del modelo (preferentemente tomando la forma de un escalar) construido según el criterio a de la sección 1.2.3 para que así el AS esté orientado hacia proveer información sobre el verdadero propósito de la construcción del modelo ya sea cualitativo, como la aceptación o rechazo de una hipótesis, o cuantitativo, como predecir la trayectoria de una partícula.

Supongamos también que es posible obtener tantas estimaciones del *output* como se deseen mediante la simulación a partir de un conjunto de valores para los *inputs*. Además, asumamos que cada *input* es una magnitud escalar que tiene un valor verdadero, que está incluido en Θ , y que ese valor no es conocido debido a la ausencia de observaciones suficientes.

Generalmente son reconocidos dos tipos de AS. Por un lado, los AS “locales” que están basados en derivadas parciales de Y respecto a un parámetro en particular cuando los demás permanecen fijos en sus valores nominales $\left(\frac{\partial Y}{\partial X_i}\right)_{X_{j \neq i} = x_j^*}$ (Asprey y Macchietto, 2000; McKay, 1999; Turanyi, 1990). Por el otro, los AS “globales” (GSA) que consideran al modelo y sus respuestas dentro del espacio Θ y por lo tanto deben ser vistos como herramientas que asisten al estudio del modelo desde una perspectiva matemática más que desde la posibilidad o no de predecir una determinada respuesta (Plischke, 2010; Campolongo, 2007; Kucherenko y Shah, 2007; Schwieger, 2004).

Una explicación pormenorizada del GSA realizado en este trabajo tanto desde el punto de vista teórico como de su implementación computacional es provista en el Anexo I. En esta tesis, los GSA serán utilizados para obtener:

- “priorización de factores” (*FP*, siglas en inglés), a partir de la cual el conjunto de parámetros de los modelos es dividido en un subconjunto de parámetros más relevantes en explicar la incertidumbre del *output* y otro cuyos elementos tienen una influencia marginal. A partir de esa división los parámetros del primer subconjunto son reestimados mientras que los restantes son fijados en su valor nominal asumiendo que ellos no contribuyen significativamente a la variabilidad del *output* del modelo (Gatelli, 2009; Jacques, 2006; Kucherenko, 2009; Sobol’, 2001; 2007; Sobol’ y Kucherenko 2009; Xu y Gertner, 2007; 2008a,b).
- estimaciones de medidas de los errores estructurales y paramétricos de los modelos para su eventual elección en el Diseño de Experimentos para Optimización.

1.3. Priorización de Factores (*FP*)

La *FP* es una técnica utilizada en el direccionamiento de la investigación. En estos casos el objetivo del GSA es identificar el o los *inputs* que mayor influencia tienen sobre la incertidumbre del *output* del modelo en estudio. Es decir, se intenta determinar cuáles fuentes de incertidumbre del modelo producirían la mayor reducción en el desconocimiento de la estimación del *output* si sus valores fuesen fijados en sus valores verdaderos.

Entonces el concepto de “importancia” de un factor en este contexto está ligado a la capacidad de este para producir una reducción en la varianza del *output*. Esto no significa que al realizar una *FP* el resultado será conocer el valor verdadero de un factor relevante sino que nos asistirá a la hora de elegir de una manera más racional en un escenario de incertidumbre. Bajo el supuesto que todos los factores son susceptibles de identificación y el costo de esa identificación es el mismo para todos, la *FP* sugerirá cuál de los factores merece el mayor esfuerzo experimental teniendo en cuenta el objetivo de mejorar las estimaciones del *output* del modelo estudiado.

Esta metodología está íntimamente ligada con el Filtrado de Factores (*FF*, siglas en inglés) que es aquella mediante la cual un AS es realizado con la intención de identificar los factores cuya repercusión en la variabilidad del *output* estudiado es despreciable (Yu y Harris, 2009; Helton y Davis, 2003). Por lo tanto, esos factores pueden ser fijados en su valor nominal para así disminuir la dimensionalidad del problema.

1.4. Mapeo de Factores. Monte Carlo Filtering

El mapeo de factores (*FM*) es la exploración de la distribución de la respuesta de un modelo y es utilizado para identificar zonas de dicha distribución que pueden resultar interesantes. Un ejemplo podría ser cuando se está interesado en obtener una respuesta Y por encima o por debajo de algún valor umbral o, en general, casos en los cuales cualquier valor de la respuesta puede ser clasificado como “sí” o “no”. Este problema conduce a la necesidad de establecer las “condiciones” o “realizaciones” de los parámetros de un modelo que hacen posible que Y tome un valor dentro del conjunto de interés.

El Monte Carlo Filtering (*MCF*) es una técnica que asiste a la resolución del problema antes mencionado. En el *MCF* son realizadas simulaciones Monte Carlo con parametrizaciones aleatorias tomadas desde el espacio de incertidumbre paramétrica Θ del modelo considerado para luego “filtrar” las respuestas obtenidas. Una vez que se cuenta con los valores aceptados para la respuesta es posible recuperar las combinaciones de los parámetros que le dieron origen y de esa manera obtener un nuevo espacio de incertidumbre paramétrica *a posteriori* $\bar{\Theta} = \{\theta / \theta \text{ produce un } Y \in \mathbf{Y}^*\}$ que cumple con $\bar{\Theta} \subset \Theta$, donde \mathbf{Y}^* es el conjunto de respuestas que cumple con la condición de “filtrado”. Una representación pictórica del método es mostrada en la figura.

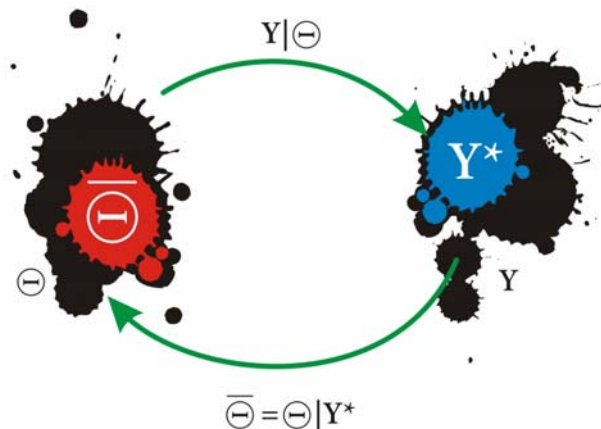


Figura 1.1: Representación de la técnica de Monte Carlo Filtering.

1.5. GSA en la práctica

1.5.1. Ejemplo de aplicación: Modelado de hidrogenación de avermectinas en medio homogéneo utilizando el catalizador de Wilkinson

El *GSA* ha sido aplicado en el desarrollo y evaluación de modelos en diferentes áreas del conocimiento (Alcázar y Ancheyta, 2007; Kiparissides, 2008; Whitcombe, 2006). En nuestra reciente contribución, “*Finding the simplest mechanistic kinetic model describing the homogeneous catalytic hydrogenation of Avermectin to Ivermectin*”, el *GSA* ha sido integrado a una metodología para el desarrollo de un modelo útil en el marco de una aplicación de actual interés: el proceso industrial catalítico más usado para la síntesis de Ivermectina (Cristaldi, 2011). Desde que este agente antihelmíntico de amplio espectro se convirtió en genérico, su precio comenzó a declinar debido al creciente aumento de la competencia, pero el precio del catalizador utilizado $[\text{RhCl}(\text{PPh}_3)_3]$ continuó acompañando la suba del metal precioso. En este escenario, la optimización del proceso de producción ha ganado una importancia creciente. De ahí la necesidad de poder contar con modelos sencillos de fácil resolución numérica para ser integrados a algoritmos rápidos de optimización y/o control de proceso. Una comparación de las predicciones del modelo más sencillo con datos experimentales tanto de un mini-reactor como de un reactor industrial operando bajo condiciones isotérmicas y no-isotérmicas, respectivamente, probó la utilidad de *GSA* para dicho propósito. Detalles pueden encontrarse en el Anexo 1.A.

A la luz de estos resultados satisfactorios y atendiendo a que la situación de tener que afrontar la resolución de problemáticas similares a la arriba descrita es frecuente en los sectores de avanzada de la Química Fina, pues éstos usualmente involucran sistemas complejos de reacciones químicas, vale en esta sección describir los principales aspectos metodológicos desarrollados como aporte de esta tesis para el modelado de sistemas complejos de reacciones químicas que pueden ser descritos por varios modelos cinéticos alternativos.

1.5.2. Metodología propuesta para el modelado de reacciones químicas

La metodología está basada en la reducción de la dimensionalidad de los diferentes modelos cinéticos mediante la técnica de *FP* para:

- Hacer un uso óptimo de la información contenida en datos experimentales de laboratorio obtenidos de aplicar un diseño experimental central compuesto restringido por condiciones operativas y restricciones económicas con los que fueron estimados los parámetros cinéticos.
- Seleccionar los mejores modelos en base a su capacidad de predecir las trayectorias de las variables de estado en condiciones experimentales de corroboración.
- Evaluar el modelo más simple, tanto desde el punto de vista estructural como paramétrico, para su utilización en la simulación de reactores de mayor porte.

Una ilustración del método es dada en la Figura 1.2. Como fue mencionado, un diseño experimental central compuesto es definido considerando las restricciones del sistema que pueden ser tanto operativas como económicas. Una vez realizadas las experiencias y recolectados los datos experimentales, estos son divididos en dos grupos: *i*) los que corresponden a las condiciones axiales del diseño experimental; *ii*) los obtenidos aplicando las condiciones centrales. El primer grupo es utilizado para la estimación ordinaria de los parámetros cinéticos de los modelos alternativos estudiados. Estos, una vez parametrizados, son sometidos a un *GSA* para conocer la relevancia de cada parámetro y realizar una *FP*. Los parámetros más importantes en la explicación de la varianza de la función objetivo (en este caso la suma de errores cuadráticos ponderada por el número de parámetros del modelo considerado) son re-estimados utilizando los datos experimentales del segundo grupo. Luego de esa parametrización más “refinada”, los intervalos de confianza de los parámetros de los modelos son calculados utilizando la

técnica *MCF*. Por último, la discriminación y selección de el/los mejor/es modelo/s es realizada en base a la suma de la función objetivo en todas las experiencias del diseño.

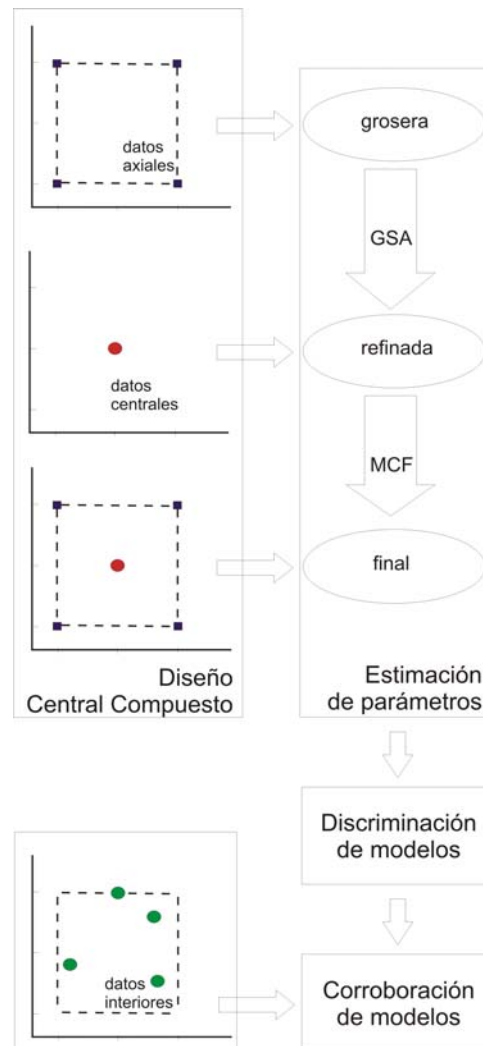


Figura 1.2: Secuencia propuesta para la selección de modelos.

1.6. Conclusiones

El esquema propuesto es de aplicación general y ha demostrado su utilidad para la discriminación entre modelos alternativos que describen sistemas complejos de reacciones químicas cuya descripción matemática involucra sistemas no lineales de ecuaciones diferenciales ordinarias y algebraicas.

La integración del *GSA* al proceso de modelado con la posterior *FP* ha asistido a la identificación de aquellos parámetros que más afectan a la variabilidad de las estimaciones junto con la estructura de interacción paramétrica.

La información suministrada por el *GSA* también ha demostrado ser útil para discriminar reacciones elementales importantes de otras que pueden ser consideradas como innecesarias para la descripción del conjunto de datos experimentales dentro de la ventana operativa en la que el proceso ha sido definido por el diseño experimental. Esto es valioso en el sentido que pueden formularse hipótesis sobre el/los mecanismo/s más probable(s) por el que transcurre la reacción considerada, evitando el uso de modelos sobrep parametrizados. En especial, vale destacar que para el sistema estudiado se ha encontrado una absoluta coincidencia entre los caminos de reacción más relevantes dentro de la intrincada red de reacciones según *GSA* y aquellos otrora sugeridos en trabajos pioneros y fundamentales que involucraron cuidadosos estudios cinéticos y complejas determinaciones espectrométricas.

Los modelos seleccionados han sido corroborados para la estimación de las trayectorias de las variables de estado bajo otras condiciones dentro del espacio experimental del diseño central compuesto (capacidad de interpolación). Además, el modelo más simple a demostrado ser capaz de explicar datos experimentales obtenidos en condiciones externas a las incluidas en el diseño experimental (capacidad de extrapolación) tanto si es considerado un reactor a escala laboratorio isotérmico como un reactor de producción a escala industrial operado bajo condiciones no isotérmicas.

1.7. Bibliografia

Alcázar, L.; Ancheyta, J. (2007). Sensitivity analysis based methodology to estimate the best set of parameters for heterogeneous kinetic models. *Chemical Engineering Journal*, **128**, 85–93.

Asprey, S.; Macchietto, S. (2000). Statistical tools for optimal dynamic model building. *Computers and Chemical Engineering*, **24**, 1261-1267.

Borgonovo, E.; Apostolakis, G. E.; Tarantola, S.; Saltelli, A. (2003). Comparison of global sensitivity analysis techniques and importance measures in PSA. *Reliability Engineering and System Safety*, **79**(2), 175-185.

Campolongo, F.; Cariboni, J.; Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling and Software*, **22**, 1509-1518.

Cristaldi, M.D.; Cabrera, M.I.; Martínez, E.C.; Grau, R.J.A. (2011). Finding the simplest mechanistic kinetic model describing the homogeneous catalytic hydrogenation of Avermectin to Ivermectin. *Industrial and Engineering Chemistry Research*. Manuscript ID ie-2010-01289h.R1 (aceptado).

Gatelli, D.; S. Kucherenko, Ratto, M., Pagano, A., Tarantola, S. (2009). Calculating first-order sensitivity measures: A benchmark of some recent methodologies. *Reliability Engineering and System Safety*, **94**(7), 1212-1219.

Georgakis, G. (1995). Modern tools of process control: The case of black, gray and white models. *Entropie*, 34–48.

Helton, J.; Davis, F. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering and System Safety*, **81**, 23–69.

Jacques, J.; Lavergne, C.; Devictor, N. (2006). "Sensitivity analysis in presence of model uncertainty and correlated inputs. *Reliability Engineering and System Safety*, **91**(10-11), 1126-1134.

Kiparissides, A.; Rodriguez-Fernandez, M.; Kucherenko, S.; Mantalaris, A.; Pistikopoulos, E. (2008). Application of Global Sensitivity Analysis to Biological Models. *ESCAPE 18*, 689-694.

Kucherenko, S.; Rodriguez-Fernandez, M.; Pantelides, C.; Shah, N. (2009); Monte Carlo evaluation of derivative-based global sensitivity measures. *Reliability Engineering and System Safety*, **94**, 1135–1148.

Kucherenko S.; Shah, N. (2007); The Importance of being Global. Application of Global Sensitivity Analysis in Monte Carlo Option Pricing; *WILMOTT Magazine*.

Leamer, E. (1990). Let's take the con out of econometrics, and Sensitivity analysis would help. Granger, C. (ed.); *Modelling Economic Series*; Oxford: Clarendon Press.

Martínez, E.C.; Cristaldi, M.D.; Grau, R.J.A. (2009). Design of Dynamic Experiments in Modeling for Optimization of Batch Processes. *Industrial and Engineering Chemistry Research*, **48**(7), 3453–3465.

McKay M.; Morrison J.; Upton S. (1999). Evaluating prediction uncertainty in simulation models. *Computer Physics Communications*, **117**, 44-51.

- Oreskes, N.; Shrader-Frechette, K.; Belitz, K. (1994). Verification, validation and confirmation of numerical model in the earth sciences. *Science*; **263**; 641-646.
- Plischke, E. (2010). An effective algorithm for computing global sensitivity indices. *Reliability Engineering and System Safety*, **95**(4), 354-360.
- Saisana, M.; Saltelli, A.; Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society A*, **168**(2),1-17.
- Saltelli, A. (2002). Sensitivity Analysis for Importance Assessment. *Risk Analysis*; **22**; 3.
- Saltelli, A.; Tarantola, S.; Campolongo, F.; Ratto, M. (2004). *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. John Wiley & Sons Ltd; England.
- Saltelli, A.; Ratto, M.; Tarantola, S.; Campolongo, F. (2006). Sensitivity analysis practices: Strategies for model-based inference. *Reliability Engineering and System Safety*; **91**, 1109–1125.
- Saltelli, M.R.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Tarantola, S. (2008). *Global Sensitivity Analysis. The Primer*. Southern Gate, Chichester, West Sussex PO19 8SQ, England, John Wiley & Sons Ltd, The Atrium.
- Schwieger V. (2004). Variance-based Sensitivity Analysis for Model Evaluation in Engineering Surveys. INGEO 2004 and FIG Regional Central and Eastern European Conference on Engineering Surveying, Bratislava, Slovakia, November 11-13.
- Sobol' I. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*; **55**, 271–280.
- Sobol, I. M.; Tarantola, S.; Gatelli, D.; Kucherenko, S.; Mauntz, W. (2007). Estimating the approximation error when fixing unessential factors in global sensitivity analysis. *Reliability Engineering and System Safety*, **92**(7), 957-960.
- Sobol', I.; Kucherenko, S. (2009). Derivative based global sensitivity measures and their link with global sensitivity indices. *Mathematics and Computers in Simulation*, **79**; 3009–3017.
- Stirling A. (2000). Risk at a turning point?. *Journal of Environmental Medicine*; **1**(3), 119-126.
- STOWA/RIZA, (1999), Smooth Modelling in Water Management, Good Modelling Practice Handbook; STOWA report 99-05, Dutch Dept. of Public Works, Institute for Inland Water Management and Waste Water Treatment, report 99.036, ISBN 90-5773-056-1.
- Turanyi, T. (1990) Sensitivity analysis of complex kinetic systems. Tools and applications. *Journal of Mathematical Chemistry*, **5**, 203–248.
- Whitcombe, J.; Cropp, R.; Braddock, R.; Agranovski, I. (2006). The use of sensitivity analysis and genetic algorithms for the management of catalyst emissions from oil refineries. *Mathematical and Computer Modelling*, **44**, 430–438.
- Xu, C.; Gertner, G. (2007). Extending a global sensitivity analysis technique to models with correlated parameters. *Computational Statistics and Data Analysis*, **51**, 5579 – 5590.

Xu, C.; Gertner, G. (2008a). A general first-order global sensitivity analysis method. *Reliability Engineering and System Safety*, **93**, 1060–1071.

Xu, C.; Gertner, G. (2008b). Uncertainty and sensitivity analysis for models with correlated parameters. *Reliability Engineering and System Safety*. **93**; 2008b,1563–1573.

Yu, W.; Harris, T. (2009). Parameter uncertainty effects on variance-based sensitivity analysis. *Reliability Engineering and System Safety*, **94**, 596–603.

*“...yo no considero que tenga muchas luces,
ni siquiera que sea inteligente,
lo que sí creo es que soy una persona tenaz,
y eso también cuenta...”*

Ernesto

Capítulo 2

Modelado para Optimización

2.1 Hipótesis de Trabajo

Durante el desarrollo y escalado de procesos químicos y bioprocesos, un diseño óptimo de experimentos desde la perspectiva del modelado para optimización permite generar incrementalmente los datos necesarios para determinar una política óptima de operación usando modelos de tendencias.

2.1.1. Modelos de Tendencias. Concepto

Los modelos de tendencia son entendidos y utilizados en este trabajo de tesis como modelos de bajo orden, estructuralmente simples que intentan dar cuenta de la estequiometría y relaciones cinéticas del sistema estudiado. Sin embargo, a pesar de su simpleza, logran describir de una forma más o menos aproximada la dinámica general del proceso. Para lo cual utilizan tanto la información disponible sobre el mismo como principios fundamentales que supuestamente gobiernan el comportamiento observado.

Los modelos de tendencias integrados en un esquema de optimización pueden ser útiles para definir una dirección de búsqueda hacia el óptimo del proceso. Sin embargo, no deben olvidarse las limitaciones de éstos para estimar con exactitud trayectorias en condiciones alejadas de las utilizadas en su implementación.

2.1.2. Modelado para Optimización. Concepto

El modelado para optimización ha surgido con la necesidad del rápido cambio de escala de procesos, principalmente los encontrados en industrias de alta tecnología, para los cuales, sea por razones de índole económica y/o de tiempo, se dispone de poco margen para realizar los extensos estudios experimentales que exige el desarrollo de modelos detallados del sistema considerado. Un buen ejemplo es la producción de una proteína en un cultivo de células de mamíferos en los que una experiencia puede tomar varios meses. Por las razones mencionadas, el paso del laboratorio a la escala piloto o de producción suele ser realizado con un alto grado de incertidumbre. Cuando se piensa en el modelado para optimización se debe abandonar la idea clásica en la cual la tarea de modelado está desacoplada y es independiente del uso futuro de esa herramienta, que en algunos casos puede ser la de optimizar un proceso.

Siguiendo la idea del modelado clásico, cada experimento es diseñado para mejorar la precisión paramétrica de los modelos con la intención de obtener de un modelo

detallado, de base mecánica mucho mejor, que sirva para la estimación del comportamiento del sistema en un rango amplio de condiciones operativas, tal como está ilustrado en la Figura 2.1.a. Sin embargo, en los casos en los que existiese un error estructural, como ocurre frecuentemente en los modelos de procesos reales, la búsqueda de una mejor precisión podría ser inútil y resultar en un gasto experimental excesivo sin mejores resultados (Srinivadan y Bonvin, 2003; Bryds y Roberts, 1987).

Bajo la modalidad del modelado para optimización, el desarrollo del modelo está integrado al algoritmo de optimización para lograr una mejora simultánea tanto del modelo como del proceso, como está ilustrado en la Figura 2.1.a. Un buen modelo es entendido como uno que sea útil para mejorar la *performance* del proceso estudiado. Esto es realizado mediante la utilización de nuevos datos experimentales obtenidos a partir de experimentos sesgados hacia la zona de operación óptima. Precisamente, una de las principales cuestiones que caracterizan el modelado para optimización es la posibilidad de diseñar experimentos que sean valiosos para el desarrollo de modelos y que a la vez asistan a la optimización de manera de sesgar iterativa y experimentalmente el modelo hacia una estructura y parametrización que conduzcan la política de operación del proceso hacia la óptima.

En el modelado para optimización, los datos recolectados deben generar información útil para reducir la incertidumbre sobre la mejor forma de operar el proceso. Eso es logrado a través de la mejora continua del modelo utilizado para el diseño de los experimentos para optimización.



Figura 2.1: Representaciones de algoritmos para el desarrollo de modelos: (a) esquema clásico y (b) modelado para optimización.

Al incluir modelos de tendencias en el esquema del modelado para optimización estarán presentes las incertidumbres tanto estructural como paramétrica. El problema se plantea

en cómo cuantificar ambas incertidumbres para realizar comparaciones, sea del aporte relativo de cada una de ellas en la calidad final de un modelo dado, sea de la capacidad de modelos diferentes para describir el sistema estudiado. Particularmente, otro de los aportes de esta tesis está vinculado a que han sido formulados indicadores que pueden ser utilizados como verdaderas medidas del error tanto estructural como paramétrico de un modelo dado.

2.1.3. Errores estructural, paramétrico y total de los modelos

El modelo para describir la evolución temporal de un proceso, según ha sido presentado en el Capítulo 1, puede ser matemáticamente representado como:

$$\frac{dx}{dt} = f(x(t), \wp(w, t), \theta, t) \quad 0 \leq t \leq t_f, \quad x(0): \text{dado} \quad (2.1)$$

donde $x(t)$ es un vector n_s -dimensional de variables de estado dependientes del tiempo, $w \in W$ es un vector m -dimensional de parámetros que definen la política de operación del proceso $\wp(w, t)$ dentro del espacio de valores admisibles W , $\theta \in \Theta$ es un vector p -dimensional de parámetros del modelo que están definidos en el espacio de incertidumbre paramétrico Θ , y t_f es el tiempo de duración del proceso.

La función objetivo utilizada para la búsqueda del óptimo es entonces definida como:

$$J = J(x, \wp(w, t), t) \quad (2.2)$$

y a su trayectoria la conoceremos por el conjunto de datos experimentales $Y = y_{jk}; \quad j = 1, 2, \dots, n_{NV} \quad k = 1, 2, \dots, n_{sp}$ obtenidos de una estrategia de muestreo ψ durante la implementación de la política de operación $\hat{\wp}(\hat{w}, t)$ con NV siendo el número de variables de estado medidas y n_{sp} el número de puntos de muestreo.

Por otra parte, numéricamente es posible realizar n experiencias Monte Carlo aplicando $\hat{\wp}(\hat{w}, t)$ al modelo $f(\bullet)$ parametrizado aleatoriamente en el espacio Θ de los parámetros del modelo. De esto resultarán n estimaciones de la trayectoria de la función objetivo \hat{J}_i que describen el espacio Ω de valores posibles de J .

Esa clase de análisis determina tanto los límites como la topología del espacio Ω definido en función de la variabilidad de los parámetros del modelo Θ . Además, la

repercusión de Θ sobre Ω está determinada en forma global y no local como ha sido planteado por algunos autores (Asprey 2000, 2002; Chen 2003).

Por cierto, la clásica medida de sensibilidad local $\left[\frac{\partial \hat{J}}{\partial \theta_i} \right]_{\theta_j; j \neq i}$ que resulta de la variación

individual de cada parámetro aquí es reemplazada por una medida que podría ser expresada como $\left[\frac{\Delta \hat{J}}{\Delta \Theta} \right]$ donde la variación de la función objetivo \hat{J} es determinada con

la modificación simultánea del conjunto de parámetros. Así, al explorar las repercusiones de Θ sobre Ω , las medidas del ajuste de las estimaciones de los modelos respecto a las trayectorias reales son buscadas dentro de ese espacio más realista y no sólo teniendo en cuenta la sensibilidad local de las estimaciones producto de la variación individual de cada parámetro.

En base a lo anterior, en un modelo pueden ser diferenciados dos tipos de errores que son producidos por diferentes fuentes de incertidumbre:

1. el error debido a la dispersión de las estimaciones producido por el desconocimiento del valor real de los parámetros del modelo (ver Figura 2.2-a).
2. el error de la estimación más probable respecto a la trayectoria real de la función objetivo determinada experimentalmente (ver Figura 2.2-b).

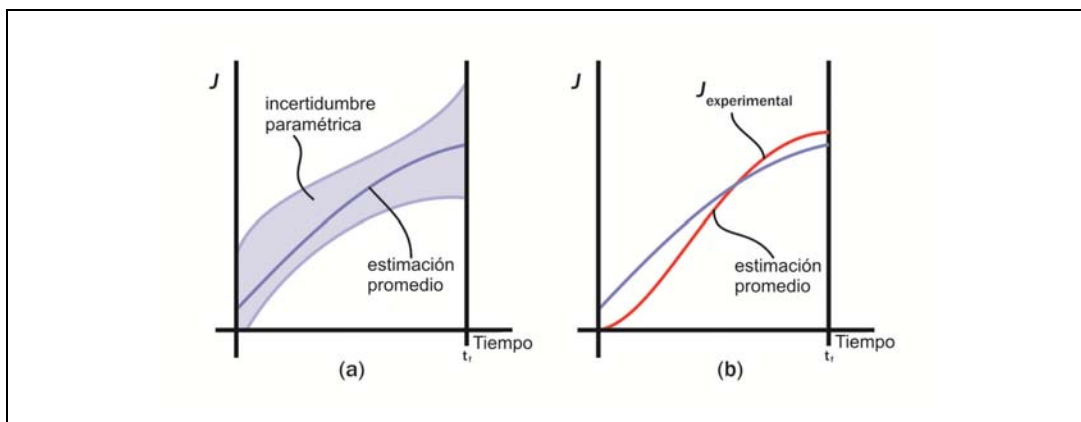


Figura 2.2: Diferentes tipos de errores. (a) incertidumbre paramétrica y (b) error estructural.

Idealmente, podrían ser planteadas estrategias para la cuantificación de ambos errores en forma continua. Sin embargo, los casos reales pueden distar mucho de esa situación debido a la imposibilidad tanto técnica como económica de contar con datos

experimentales de la función objetivo en todo el intervalo de tiempo durante el cual el proceso tiene lugar. Es por eso que, en adelante, aquí serán planteadas estrategias de estimación de ambos tipos de errores con formulaciones discretas, las cuales pueden ser fácilmente extendidas para incluir el caso continuo ideal.

La definición precisa de la función objetivo J , cuya expresión matemática general es dada por la Ec. 2.2, no es trivial. El experimentador debe conciliar la especificación del objetivo buscado con una ecuación matemática que exprese exactamente sus deseos. Esto es esencial ya que la metodología de discriminación de modelos y optimización se encuentra basada en dicha definición, y por lo tanto los resultados de su implementación dependerán en forma directa de la misma.

Error paramétrico

El error paramétrico es definido como:

$$\varepsilon_{\hat{J}_i \rightarrow \langle \hat{J}_i \rangle} = \frac{1}{n_{sp} n} tr \left[\left(\hat{J}_{ij} - \langle \hat{J}_i \rangle \right) W_{ij} \left(\hat{J}_{ij} - \langle \hat{J}_i \rangle \right)^T \right]; \quad i = 1, 2, \dots, n_{sp}; j = 1, 2, \dots, n \quad (2.3)$$

donde \hat{J}_{ij} es la estimación de la función objetivo en el i -ésimo tiempo de muestreo para la j -ésima simulación utilizando el modelo estudiado; $\langle \hat{J}_i \rangle$ es la estimación más probable de la función objetivo en el i -ésimo tiempo de muestreo; y W_{ij} es una matriz que pondera la importancia relativa de las estimaciones en los diferentes tiempos de muestreo en base a las creencias o subjetividades del experimentador.

Esta formulación intenta cuantificar el error producido principalmente por la falta de conocimiento sobre el valor real de los parámetros del modelo $\left[\hat{J}_{ij} - \langle \hat{J}_i \rangle \right]$ y la dispersión producida debido a Θ .

Como el indicador $\varepsilon_{\hat{J}_i \rightarrow \langle \hat{J}_i \rangle}$ es estimado en un entorno de sensibilidad global se sobrepasa el problema de normalización de las estimaciones de sensibilidad cuya perspectiva es básicamente local. Por otro lado, los problemas de adjudicarle mayor relevancia a aquellos parámetros que se conocen con mayor incertidumbre son evitados ya que la influencia de cada parámetro individual estará ponderada en un entorno de variabilidad paramétrica conjunta. Con esta formulación lo que se intenta es medir la

convergencia puntual de las estimaciones de la función objetivo respecto a la trayectoria más probable utilizando el modelo cuyo error se intenta conocer.

Por cuestiones de simplicidad y homogeneidad se plantea considerar los tiempos de muestreo como puntos de estimación del error paramétrico $\varepsilon_{J_i \rightarrow \langle \hat{J}_i \rangle}$. Sin embargo, esta no es una restricción ya que las simulaciones posibilitan contar con mayor cantidad de puntos los cuales pueden ser considerados al momento de estimar dicho error.

Error estructural

El error estructural es definido como:

$$\varepsilon_{\langle J \rangle \rightarrow J} = \frac{1}{n_{sp}} \left[\left(\langle \hat{J}_i \rangle - J_i \right)^T W_{ii} \left(\langle \hat{J}_i \rangle - J_i \right) \right]; \quad i = 1, 2, \dots, n_{sp} \quad (2.4)$$

donde J_i es la medida experimental de la función objetivo en el i -ésimo tiempo de muestreo; y W_{ii} es una matriz que pondera la importancia relativa de las estimaciones de la función objetivo en los diferentes tiempos de muestreo en base a las creencias o subjetividades del experimentador.

Con esta formulación se intenta cuantificar principalmente la incertidumbre introducida por la estructura (forma funcional) del modelo y mide la distancia entre la trayectoria más probable estimada y el valor real de la función objetivo.

Con esta definición nuevamente se busca tener una medida de la convergencia puntual de las estimaciones respecto al valor real de la función objetivo J para todo el intervalo de tiempo en el que el proceso se lleva a cabo.

Error total del modelo

El error total de las estimaciones de J es definido como:

$$E'_J = \lambda_{\langle J \rangle \rightarrow J} \varepsilon_{\langle J \rangle \rightarrow J} + \lambda_{J_i \rightarrow \langle \hat{J}_i \rangle} \varepsilon_{J_i \rightarrow \langle \hat{J}_i \rangle} \quad (2.5)$$

donde $\varepsilon_{J_i \rightarrow \langle \hat{J}_i \rangle}$ es el error asociado a la dispersión de las estimaciones respecto a la trayectoria estimada más probable, es decir el error paramétrico; y $\varepsilon_{\langle J \rangle \rightarrow J}$ es el error asociado a la distancia entre la estimación más probable de la evolución del sistema y la trayectoria real de la función objetivo, o error estructural. En esta formulación general, los λ son factores de peso que el experimentador puede manipular según lo considere para el caso de estudio particular.

Esta división de las fuentes de incertidumbre respecto del error total intenta desacoplar las contribuciones de las incertidumbres paramétrica y estructural de un modelo, y asociarlas a las dispersiones tanto de las estimaciones Ω como de la trayectoria real J , respectivamente.

Medidas independientes de los errores tanto estructural como paramétrico de un modelo permiten establecer prioridades en la investigación ya que:

- Un error estructural grande respecto al paramétrico estaría evidenciando un problema en la formulación matemática o bien una mala definición de Θ como resultado de la importación de valores o intervalos de parámetros inadecuados obtenidos por experimentos mal realizados, o de casos de sistemas que se presuponen similares pero que en verdad no se ajustan al proceso estudiado.
- Un error paramétrico grande respecto al estructural podría interpretarse como una falta de información del sistema. El modelo, en principio, sería correcto ya que, en términos medios, es capaz de capturar la dinámica del proceso. La incertidumbre en las estimaciones del modelo pueden mejorarse disminuyendo el desconocimiento en los parámetros, lo cual es posible de hacer contando con una mayor cantidad de datos experimentales.
- Ante el caso de una multiplicidad de modelos para la descripción de un mismo sistema, tanto el error total como los aportes desagregados contribuirían a la elección del más pertinente ya que el método tiene en cuenta la capacidad de estimación más probable a través de $\varepsilon_{(J) \rightarrow J}$ y la dispersión total de las estimaciones a través de los espacios de incertidumbre paramétrica, sin focalizar únicamente en el número de parámetros incluidos en cada modelo.

Criterio de selección del “mejor” modelo

Cuando se cuenta con más de un modelo como herramienta de predicción, la definición del “mejor” no constituye una trivialidad. En esta tesis la discriminación entre modelos alternativos no pretende ser absoluta sino que se propone la selección de aquel que presenta el menor valor del error total en cada iteración calculado según ha sido explicado anteriormente en la sección 2.1.3.

Definiendo E_7^g el error total para el g-ésimo modelo como:

$$E_T^g = \lambda_{\langle \hat{j} \rangle \rightarrow J}^g \varepsilon_{\langle \hat{j} \rangle \rightarrow J}^g + \lambda_{J_i \rightarrow \langle J \rangle}^g \varepsilon_{J_i \rightarrow \langle J \rangle}^g; \quad g = 1, 2, \dots, k \quad (2.6)$$

la medida del error de cada modelo, entre los k disponibles, tiene en cuenta su desempeño desde una visión global que permite ponderar diferentes aspectos que hacen de un modelo una buena, regular o mala herramienta para optimización (certeza paramétrica y estructural).

Precisamente, en el error total se computan los aportes a la incertidumbre desde:

- la estructura del modelo a través de $\varepsilon_{\langle J \rangle \rightarrow J}$
- los parámetros del modelo a través de $\varepsilon_{\hat{j}_i \rightarrow \langle \hat{j}_i \rangle}$

Entonces, el mejor modelo será el que cumpla con:

$$E_T^{best} = \min_g E_T^g; \quad g = 1, 2, \dots, k \quad (2.7)$$

2.2. Diseño de Experimentos Dinámicos para Optimización basado en Modelos

La mayoría de las técnicas de optimización utilizan modelos del sistema en estudio. Sin embargo, no es frecuente contar con un modelo confiable para operar en forma óptima un proceso. En general, existen modelos aproximados en los que cualquier *mismatch* entre el proceso y el modelo es compensado mediante el ajuste de los parámetros del modelo utilizando datos experimentales.

En el Modelado para Optimización la idea subyacente es que luego de cada experimento, producto de aplicar políticas de operación diseñadas para mejorar la *performance* del proceso, los parámetros del modelo sean sesgados para una mejor estimación de la política de operación óptima. Así, el modelado no es un fin en si mismo sino sólo un medio o herramienta que guía la búsqueda de la mejor forma de operar el sistema con el mínimo de esfuerzo experimental.

Teniendo en cuenta que la obtención de medidas experimentales puede ser costosa, la mejora del modelo para optimización es producto de la reducción en la incertidumbre paramétrica mediante la re-estimación sólo de los parámetros que tienen mayor efecto sobre la estimación de la *performance* del proceso. Esa relevancia es determinada mediante un *GSA* del modelo operando en las condiciones actuales y luego de una *FP* para así poder identificar aquellos parámetros que persistentemente excitan a la función objetivo del modelo (Srinivadan y Bonvin, 2003).

2.2.1. Metodología propuesta

La metodología de optimización propuesta es experimental e iterativa. Ella está basada en la utilización de modelos de tendencias para el diseño de experimentos dinámicos óptimos.

Esos experimentos son entendidos como:

- un conjunto de condiciones operativas $w \in \mathcal{W}$ que hacen óptimo el valor de una función objetivo J .
- una estrategia de muestreo $\psi \in \Psi$ óptima en el sentido que los tiempos a los que se deben tomar las muestras corresponden a momentos en los cuales los parámetros que definen la política de operación producen una excitación máxima de J .

En la Figura 2.3a es mostrado un esquema del método iterativo para optimización propuesto. El caso general es aquel en el cual existe una biblioteca de modelos de tendencias que puede ser utilizada para la búsqueda del óptimo.

En cada iteración una política de operación óptima es implementada en el proceso y un nuevo conjunto de datos es obtenido a partir de un muestreo óptimo. Esos datos son utilizados para la re-estimación de los parámetros más inciertos de cada modelo en la biblioteca según el *ranking* obtenido del *GSA*.

Los errores estructural, paramétrico y total son calculados, y a partir de ellos el mejor modelo es seleccionado para calcular el próximo experimento.

Para el caso ideal en el que la biblioteca de modelos cuenta con uno que es estructuralmente correcto y, además, no existen errores de medición en los datos experimentales, la metodología converge a la política óptima del proceso en una única iteración.

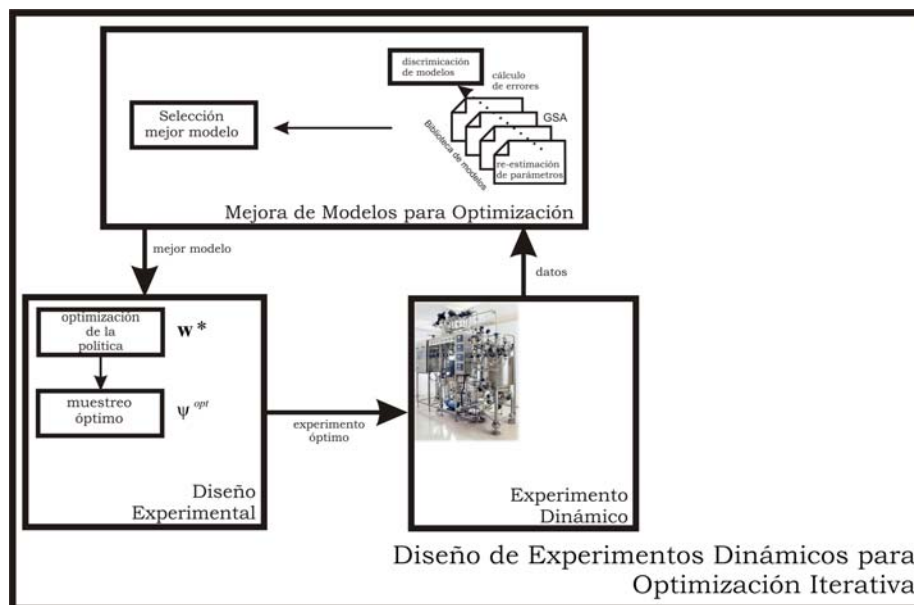


Figura 2.3a: Diseño de experimentos dinámicos para optimización iterativa.

2.2.1.1. Iteración de la política de operación basada en modelos

Nuevamente, se considera que el proceso a optimizar puede ser representado por un sistema de ecuaciones diferenciales ordinarias:

$$\frac{dx}{dt} = f(x(t), \varphi(w, t), \theta, t) \quad 0 \leq t \leq t_f, \quad x(0) : \text{dado} \quad (2.8)$$

y además que la función objetivo a ser maximizada para alcanzar el óptimo puede escribirse como:

$$J(w) = h(x(t_f)) + \int_0^{t_f} g(x(t), \wp(w, t), \theta, t) dt \quad (2.9)$$

donde $x(t)$ es un vector de dimensión n_s que representa a las variables de estado del sistema; $w \in W$ es un vector m -dimensional de parámetros que definen la política de operación del proceso considerado representada por $\wp(w, t)$ y W es el espacio de valores que pueden adoptar esos parámetros ya sea porque existen restricciones físicas y/o operativas; $\theta \in \Theta$ es un vector de dimensión p correspondiente a los parámetros del modelo que pertenece al dominio de incertidumbre paramétrica Θ ; y t_f es el tiempo final del proceso. La función $g(\bullet)$ es la función de recompensa para una dada parametrización de la política $\wp(w, t)$ mientras que la función $h(\bullet)$ es la función de recompensa para el final de la corrida por haber utilizado dicha política.

Las políticas de operación $\wp(w, t)$ utilizadas cuentan de dos tipos de componentes: (i) aquellos que no se modifican durante todo el proceso (por ejemplo, condiciones iniciales, tiempo de inducción del cultivo) y (ii) los que varían en función del tiempo (por ejemplo, perfil de alimentación del biorreactor).

Para el caso de un modelo individual y una dada parametrización $\hat{\theta}$, la política de operación óptima determinada por el modelo es el resultado de resolver el problema que satisface la condición de optimalidad de *Hamilton-Jacobi-Bellman (HJB)* (Bellman, 1958):

$$\frac{\partial J(w^*)}{\partial t} = \min_{w \in W} \left\{ g(x, \wp^*, \hat{\theta}, t) + \left[\frac{\partial J(w^*)}{\partial x} \right]^T f(x, \wp^*, \hat{\theta}, t) \right\}, \quad (2.10)$$

para todo t, x ,

$$S(x, w, \hat{\theta}) \leq 0$$

$$T(x(t_f), w, \hat{\theta}) \leq 0$$

con la condición de borde $J(w^*) = h(x(t_f))$. S y T representan las restricciones operativas, económicas, u otras, que la política debe cumplir durante la trayectoria del

proceso y al final del ensayo experimental, respectivamente. Es decir que S y T las fronteras de W .

La condición de optimalidad (Ec. 2.10) es una ecuación en derivadas parciales que debe ser satisfecha por la solución $J(w^*)$ para todos los pares (t, x) .

La política de operación óptima $\varphi^*(w^*, t)$ puede ser iterativamente encontrada maximizando el lado derecho de la Ec. (2.10). Por ejemplo, una política definida por los parámetros w_2 que produce un índice de *performance* J_2 es mejor que (o preferible a) una política w_1 que produce un J_1 si y sólo si $J_2 > J_1$.

Cuando un modelo de tendencias es capaz de capturar correctamente la dinámica del proceso (error estructural nulo) y además sus parámetros son perfectamente conocidos (error paramétrico nulo), el diseño de experimentos basado en modelos conduciría a la solución óptima en la primera iteración. Sin embargo, debido a que la incertidumbre paramétrica en general es grande y las determinaciones experimentales contienen además una fuente de incertidumbre adicional debida al ruido experimental, generalmente deben ser diseñados varios experimentos óptimos antes de alcanzar la convergencia en la política de operación óptima del proceso.

El método iterativo comienza con un conjunto inicial de parámetros del(los) modelo(s) $\hat{\theta}$ los cuales, por ejemplo, pueden ser los valores medios de los intervalos de variación conocidos o los valores aportados por la opinión de expertos. Con esa parametrización grosera, una primera política de operación es obtenida como resultado de resolver el problema dado por la Ecuación (2.10). Una alternativa para comenzar el método es a partir de una política de operación basada en el conocimiento previo de operarios del proceso o en buenas prácticas. Luego, la política es sucesivamente mejorada por el proceso iterativo que aporta nueva información experimental para reducir la incertidumbre paramétrica.

En general, la política inicial suele ser conservadora y sub-óptima, no obstante puede ser utilizada desde la primera iteración para sesgar el(los) modelo(s) de tendencias hacia una parametrización para la optimización. Eso es logrado asegurando que luego de implementar una política de operación estimada, la excitación persistente de la función objetivo es producida (en su mayor parte) por un subconjunto de parámetros del modelo θ_i que contiene los más relevantes para reducir la diferencia entre la trayectoria real y la

estimada (ver sección 2.2.2). El subconjunto θ_i es determinado utilizando la técnica de *FP* luego de un *GSA*, y sus componentes son re-estimados utilizando los datos experimentales mientras los demás parámetros del modelo son mantenidos constantes en sus valores nominales.

Es considerado que el método ha convergido a la política de operación óptima si se cumple alguna de las siguientes condiciones:

$$J_{experimental}^{n-1} = J_{experimental}^n \quad (2.11)$$

$$\wp^{n-1}(w^{n-1}, t) \equiv \wp^n(w^n, t) \quad (2.12)$$

$$std[J_{experimental}] \geq std[J_{estimada}] \quad (2.13)$$

2.2.1.2. Muestreo óptimo

Como aseguran Banks y col. (Banks y col., 2009), en cualquier conjunto de datos experimentales que pueda ser tomado durante un proceso la información para la estimación de un conjunto de parámetros cinéticos puede variar considerablemente. Entonces, es posible utilizar alguna herramienta matemática que permita la recolección de datos en forma óptima para la estimación de parámetros o para resolver el problema inverso; esto es, dado un conjunto de datos experimentales poder conocer aquellos parámetros que pueden ser estimados con mayor grado de certeza (ver sección 2.2.2).

Una vez que la política de operación óptima (para una dada parametrización del modelo de tendencias en la iteración n) ha sido definida, resta calcular cuáles serían los mejores momentos para realizar el muestreo del proceso dado que no siempre es posible medir continuamente ciertas variables de estado. La estrategia de muestreo a implementar debe aportar información relevante para la mejor parametrización de las herramientas de optimización, es decir los modelos de tendencia, y por lo tanto para estimar con mayor certeza la función objetivo. Además, para una dada parametrización de la política de operación existen variaciones en la implementación que la estrategia de muestreo debería ayudar a monitorear.

Asumiendo que para la n -ésima iteración una política de operación óptima $\wp(w_n, t)$ ha sido calculada por un modelo cuya parametrización es $\hat{\theta}$, entonces el objetivo de encontrar la estrategia de muestreo óptima $\psi^{opt} \in \Psi$ es determinar aquellos momentos

en los cuales los datos obtenidos aportarán mayor información sobre el proceso. Ψ es el espacio de tiempos de muestreo posibles, es decir restringido por el número de muestras n_{sp} que se pueden (o desean) realizar durante el proceso, el mínimo y máximo tiempo intermuestra.

En esta metodología, el encontrar los tiempos de muestreo óptimo se reduce a resolver el problema:

$$\begin{aligned} \psi^{opt} &= \max_{\psi \in \Psi} \det |M|, \\ M(\hat{\theta}, \wp(w_n, t), \psi) &= Q^T Q \\ Q &= \begin{pmatrix} Si_{11} & \cdots & Si_{1n} \\ \vdots & \ddots & \vdots \\ Si_{m1} & \cdots & Si_{mn} \end{pmatrix} \end{aligned} \quad (2.14)$$

sujeto a las restricciones:

$$t_{\min}^{\text{int}} \leq (t_i - t_{i-1}) \leq t_{\max}^{\text{int}}; i = 1, \dots, n_{sp} \quad (2.15)$$

$$t_{n_{sp}} \leq t_f \quad (2.16)$$

donde cada elemento Si_{ij} de la matriz Q es una medida de la sensibilidad del índice de *performance* del proceso $J(\hat{w}_n)$ en el i -ésimo tiempo de muestreo respecto al j -ésimo parámetro de la política de operación; t_{\min}^{int} y t_{\max}^{int} son los tiempos intermuestra mínimo y máximo permitidos y t_f es el tiempo final del proceso; y $\det |M|$ es el determinante de la matriz M .

El número de muestras n_{sp} estará definido por el presupuesto disponible para cada experiencia pero con la restricción que como mínimo debe tener tantas muestras como número de parámetros que definen la política (o los parámetros inciertos dentro de la política).

2.2.2. Actualización óptima de modelos y selección del mejor modelo para optimización

En la n -ésima iteración, luego de realizar el experimento diseñado de una manera óptima que consiste tanto en una política de operación del proceso $\wp(w_n, t)$ como de

una estrategia de muestreo ψ_n , se cuenta con una matriz de datos experimentales Y_n^{exp} de (n_{sp}, n_s) .

Además, para una política $\wp(w_n, t)$ es realizado un *GSA* a $f(x(t), \wp(w_n, t), \theta, t); \theta \in \Theta_n$ respecto a su capacidad de estimar $J(w_n)$ y se obtiene una FP en la que θ es dividido en dos sub-conjuntos θ_l de parámetros relevantes y θ_{p-l} de parámetros cuyo grado de incertidumbre no influye significativamente en la estimación de $J(w_n)$ y por lo tanto pueden ser fijados en sus valores nominales (ver anexo 1.1 para más detalles).

La mejora del modelo para optimización es realizada al resolver el problema:

$$F = \min_{\theta_l \in \Theta_l} \left(\text{tr} \left[\left(\left(Y_n^{\text{exp}} - Y(\theta_{l-p}, \theta_l) \right)^2 \right)^T \left[\left(Y_n^{\text{exp}} - Y(\theta_{l-p}, \theta_l) \right)^2 \right] \right] \right) \quad (2.17)$$

donde F es la función objetivo de mejora de los modelos que, como lo han propuesto Srivadan y Bonvin (Srinivadan y Bonvin, 2003), no coincide con la de optimización; $Y(\theta_{l-p}, \theta_l)$ es la estimación hecha por el modelo parametrizado con $\theta = (\theta_{l-p}, \theta_l)$.

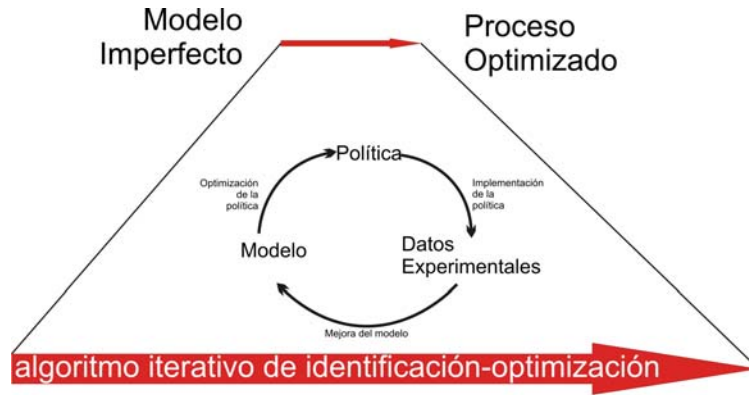
Luego es realizado un Monte Carlo Filtering para conocer los intervalos de variación de los parámetros en θ_l .

En los casos en los que existe una biblioteca de modelos para optimización, los errores paramétricos y estructurales son calculados y aquel que presenta el menor error total es escogido para comenzar una nueva iteración en el diseño de experimentos dinámicos para optimización basado en modelos.

2.2.3. Convergencia del método

El análisis de la convergencia de esquemas iterativos de identificación-optimización que utilizan modelos imperfectos ha sido tratado y resaltado en diferentes trabajos (Srinivadan y Bonvin, 2003; Brdyš y Roberts, 1987). Como esta ilustrado en la Figura 2.3b, los principales interrogantes cuando son utilizadas estas estrategias de diseño de experimentos dinámicos para optimización están relacionados a las siguientes cuestiones:

- La convergencia del algoritmo a una política cualquiera.
- Bajo qué situaciones o qué requerimientos debe cumplir el método para lograr una convergencia hacia el (o en una zona cercana al) óptimo real del proceso.



interrogantes principales: el ciclo converge a una política de operación a pesar de los errores del modelo?

la mejor política estimada por el modelo es cercana a la política de operación óptima real?

Figura 2.3b: Optimización de procesos en base a modelos imperfectos utilizando algoritmos experimentales iterativos.

Para un modelo descrito por la Ec. (2.1), cuyo índice de performance es expresado por la Ec. (2.2), la condición de optimalidad de HJB puede ser reescrita como:

$$0 = \min_{w \in H'} \left\{ g(x, \wp(w, t), \hat{\theta}, t) + \frac{\partial J(t, w)}{\partial t} + \left[\frac{\partial J(t, w)}{\partial x} \right]^T f(x, \wp(w, t), \hat{\theta}, t) \right\} \quad (2.18)$$

si en el proceso es implementada alguna política de operación sub-óptima $\wp(\hat{w}, t)$, entonces existe un error que puede ser escrito en términos de la ecuación anterior como:

$$0 < g(x, \wp(\hat{w}, t), \hat{\theta}, t) + \frac{dJ(t, \hat{w})}{dt} \quad (2.19)$$

Integrando la Ec. (2.19) entre los tiempos inicio y final del proceso estudiado es obtenido el residuo de Bellman BR :

$$BR(\hat{\theta}, \hat{w}) = \left[\int_0^{t_f} g(x(t), \wp(\hat{w}, t), \hat{\theta}, t) dt + h(x(t_f)) - J(0, x(0)) \right] > 0 \quad (2.20)$$

BR está acotado por debajo en cero y tomará dicho valor sólo cuando con un modelo, cuya estructura sea correcta y esté perfectamente parametrizado $\theta = \theta_{real}$, sea obtenida la política de operación óptima $\wp^*(w^*, t)$.

Para cualquier situación en la que una política de operación $\hat{\wp}(\hat{w}, t)$ sea calculada utilizando un modelo estructural o paramétricamente imperfecto, el BR puede ser reescrito en función del $BR(\theta_{real}, w^*) = 0$ del modelo ideal (estructuralmente correcto y perfectamente parametrizado) con el que se puede calcular la política óptima:

$$BR(\hat{\theta}, \hat{w}) - BR(\theta_{real}, w^*) = \int_0^{t_f} \left[g(x(t), \wp(\hat{w}, t), \hat{\theta}, t) - g(x_{real}(t), \wp(w^*, t), \theta_{real}, t) \right] dt \dots \quad (2.21)$$

$$\dots + [h(x(t_f)) - h(x_{real}(t_f))]$$

$$BR(\hat{\theta}, \hat{w}) - BR(\theta_{real}, w^*) = J(t_f, x) - J^*(t_f, x_{real}) \quad (2.22)$$

En base a la Ec. (2.22), la política de operación óptima $\wp^*(w^*, t)$ puede ser encontrada mediante la minimización iterativa del error entre las estimaciones de la trayectoria de J al aplicar una política optimizada con $f(x(t), \wp(w, t), \hat{\theta}, t)$ y la trayectoria real. Entonces, cada iteración comienza con una política de operación $\hat{\wp}(\hat{w}, t)$ producto del diseño de un experimento dinámico. Ella es implementada en el proceso real desde el cual son obtenidos datos experimentales $J_{exp}(t_r, x_{exp})$, $r = 1, \dots, n_{sp}$. Esos datos son utilizados para mejorar la capacidad predictiva del(los) modelo(s) y consecuentemente reducir el error en las predicciones:

$$E(\hat{\theta}) = \sum_{r=1}^{n_{sp}} \left[J(t_r, x(t, \hat{w}, \hat{\theta})) - J_{exp}(t_r, x_{exp}) \right]^2 \quad (2.23)$$

De lo anterior, aparece como fundamental el utilizar la información experimental generada en cada iteración de manera de garantizar la convergencia del método de optimización mediante la reducción del error E en la predicción del índice de *performance*.

Si $E \equiv 0$, $\hat{w}_n = w^*$ y el método ha convergido al óptimo del proceso.

Condiciones para la convergencia

Teniendo en cuenta lo anterior, las condiciones para la convergencia resultan de considerar el problema de minimizar el error entre la predicción y la trayectoria real de la función objetivo $E: \Theta \rightarrow \Re$. Ese error será una función de la parametrización del

modelo $f(\hat{\theta})$ y su correspondiente predicción de la política de operación óptima $\hat{\phi}(\hat{w}, t)$, donde Θ es el espacio de incertidumbre paramétrico que se intenta reducir en cada iteración.

Sea Γ la condición de optimalidad de *HJB* y sea $\Theta^\Gamma \subset \Theta$ el conjunto de parametrizaciones para las cuales Γ se satisface, es decir que se acepta la multiplicidad (o existencia de un subespacio) de parametrizaciones del modelo que pueden conducir al óptimo.

Además, $\xi: \Theta \rightarrow \mathfrak{R}^-$ es una función de valores no-positivos tal que $\xi \equiv 0 \Leftrightarrow \hat{\theta} \in \Theta^\Gamma$. ξ se denominada la función de optimalidad asociada a Γ (Polak, 1997) y debe estar definida de forma que otorgue medidas cuantitativas de cuan cerca o lejos está una parametrización $\bar{\theta}$ y la correspondiente política optimizada \bar{w} de la condición Γ , es decir si $\bar{\theta} \in \Theta^\Gamma$.

Teniendo en cuenta lo anterior, la función de optimalidad para el análisis de convergencia de un algoritmo de optimización surge naturalmente, es el módulo de *BR*:

$$\xi(\bar{\theta}, \bar{w}) = -|BR(\bar{\theta}, \bar{w})|, \bar{\theta} \in \Theta, \bar{w} \in W \quad (2.24)$$

Definición 1. El sumidero de Θ

A medida que el número de iteraciones n aumenta, el espacio de incertidumbre paramétrica Θ presenta un comportamiento de estrechamiento (no necesariamente monótono) hacia un punto de acumulación o sumidero $\tilde{\theta}$. Matemáticamente:

$$\lim_{n \rightarrow \infty} \Theta^n = \tilde{\theta} \quad (2.25)$$

donde el espacio paramétrico que cumple con la condición de poseer un punto de acumulación para algún n finito satisface que $\Theta^{n+1} \subset \Theta^n$.

Definición 2. Descenso suficiente del algoritmo de identificación-optimización

Un algoritmo de mejora de la política de operación de un proceso basado en modelos que presenta un sumidero en el dominio Θ , además tiene la propiedad de descenso suficiente respecto a la función de optimalidad ξ si para cada $\delta > 0$ existe un $\eta < 0$, donde para cada i -ésima iteración:

$$\xi(\hat{\theta}^i) < -\delta \quad (2.26)$$

$$E(\hat{\theta}^{i+1}) - E(\hat{\theta}^i) < \eta \quad (2.27)$$

donde $\xi(\hat{\theta}^i)$ sería el residuo de Bellman dada la parametrización $\hat{\theta}^i$ y es el la diferencia entre las trayectorias estimada y la real de la función objetivo para la $E(\hat{\theta}^i)$ j -ésima iteración.

Propuesta

Suponiendo que existe una constante $D \in \mathfrak{R}$ tal que $E(\theta) \leq D, \forall \theta \in \Theta$, si un algoritmo iterativo para la optimización de la política de iteración de un proceso basado en modelos tiene la propiedad de descenso suficiente, entonces, para cualquier secuencia infinita $\{\hat{\theta}^i\}_{i=1}^{\infty}$ se cumple que:

$$\lim_{i \rightarrow \infty} \xi(\hat{\theta}^i) = 0 \quad (2.28)$$

La prueba es inmediata desde la definición 2.

Según las definiciones y la propuesta anteriores, para garantizar que un algoritmo iterativo para optimización basado en modelos funciona, es necesario que la incertidumbre paramétrica del modelo presente un punto de acumulación.

En los ejemplos que siguen esa condición ha sido impuesta mediante la definición del dominio de incertidumbre paramétrica introducido en el *MCF* luego de cada iteración.

2.3. Ejemplo 1: Diseño de Experimentos Dinámicos para la Optimización de la producción de penicilina en un cultivo de *Penicillium crysogenum*. Proceso idéntico al modelo.

Debido a la larga historia de producción de penicilina y el volumen de conocimiento generado en torno al proceso de fermentación que involucra su obtención, este sistema ha sido establecido como punto de referencia para la corroboración de nuevas herramientas de modelado, optimización, monitoreo y control de bioprocesos novedosos. Por lo tanto, este bioproceso es inicialmente adoptado para ilustrar la aplicación de la metodología de optimización.

Industrialmente, la penicilina es producida en cultivos de *Penicillium crysogenum* en los que, luego de una primera etapa batch para favorecer el crecimiento celular, el cultivo es llevado a una situación de estrés metabólico logrando así que las células produzcan y liberen el antibiótico al medio de cultivo. Generalmente, esta última etapa del cultivo es realizada en modo *fed-batch* pues va acompañada de la adición continua de la fuente de carbono necesaria para el mantenimiento celular, junto con un inductor para la producción de alguna penicilina específica.

En términos generales el proceso puede simplificarse a la obtención de biomasa X y penicilina P a partir de un sustrato S , generalmente glucosa, limitante del desarrollo celular y de alguna fuente de nitrógeno, tal como el *corn steep licour*. Las variables que comúnmente son utilizadas para el modelado de la dinámica del proceso son las que rutinariamente son medidas como:

- concentración de células viables X_v y muertas X_d
- concentración del sustrato limitante S
- producción de penicilina P
- volumen del cultivo V

El cultivo es realizado en forma *fed-batch* ya que con cierta frecuencia es necesario producir descargas de un volumen de medio de cultivo para mantener tanto la viabilidad celular como el cultivo anéxico del hongo productor de penicilina evitando así posibles mutaciones y pérdida de productividad de la cepa de *P. crysogenum*.

En la práctica industrial, la política de operación del proceso es establecida para tratar de maximizar la cantidad de penicilina obtenida al final del cultivo.

La metodología propuesta es utilizada para la optimización de la producción de penicilina en un biorreactor *in silico*. Tanto la estructura matemática del modelo como Θ han sido tomadas del trabajo de Menezes y col. (1994). En concordancia, el modelo matemático es descrito por el siguiente sistema de ecuaciones diferenciales:

$$\begin{aligned}
\frac{dV}{dt} &= F_{in} - F_{evap} \\
\frac{dX_v}{dt} &= \mu X_v - K_d X_v - \frac{(F_{in} - F_{evap})}{V} X_v; \quad \mu = \frac{\mu_{m\acute{a}x} S}{K_x X_v + S} \\
\frac{dX_d}{dt} &= K_d X_v - k_{lis} X_d - \frac{(F_{in} - F_{evap})}{V} X_d \\
\frac{dS}{dt} &= -\sigma X_v + \langle S_{in} F_{in} \rangle - \frac{(F_{in} - F_{evap})}{V} S; \quad \sigma = \frac{\mu}{Y_{xs}} + \frac{\pi}{Y_{ps}} + \zeta; \quad \zeta = \frac{\zeta_{m\acute{a}x} S}{K_s + S} \\
\frac{dP}{dt} &= \pi X_v - \frac{(F_{in} - F_{evap})}{V} P - K_h P; \quad \pi = \frac{\pi_{m\acute{a}x} S}{K_p + S}
\end{aligned} \tag{2.3.1}$$

sujeto a las siguientes condiciones iniciales:

$$X(0) = 1; S(0) = 4; P(0) = 0$$

donde F_{in} es el caudal de ingreso de nutrientes, F_{evap} es la pérdida de líquido debido a la evaporación del medio de cultivo y S_{in} es la concentración de sustrato en la alimentación del biorreactor.

El modelo es estructuralmente correcto y la parametrización inicial no se corresponde con la del biorreactor real pero está incluida dentro del espacio de incertidumbre paramétrica, es decir que $\theta_{real} \in \Theta$.

La parametrización inicial del modelo es realizada adoptando los valores medios de los intervalos reportados por Menezes y col., tal como se muestra en la Tabla 2.3.1. De este modo, la convergencia del método es estudiada para el caso más simple donde sólo existe incertidumbre paramétrica y ruido experimental (10% del valor de las variables de estado individuales).

Particularmente, la capacidad y utilidad del método es examinada para cuatro perfiles distintos de alimentación del biorreactor. En todos los casos la estrategia de muestreo inicial es equidistante con una muestra cada 24 hs.

Tabla 2.3. 1: Intervalos de incertidumbre del modelo de Menezes (Menezes y col., 1994) y parametrización del biorreactor *in silico*

Parámetros	Símbolo	Unidades	Rango inicial	biorreactor <i>in silico</i>
<i>Velocidad máxima de crecimiento específica</i>	$\mu_{m\acute{a}x}$	h^{-1}	0,12 – 0,17	0,13
<i>Constante de saturación del crecimiento celular</i>	K_x	g-sustrato/g-DW	0,006 – 0,400	0,131
<i>Velocidad de muerte celular</i>	K_d	h^{-1}	0,005 – 0,010	0,006
<i>Velocidad de lisis celular</i>	k_{lis}	h^{-1}	0,00001 – 0,00800	0,0008
<i>Rendimiento de biomasa respecto de sustrato</i>	Y_{xs}	g-DW/g-sustrato	0,40 – 0,58	0,52
<i>Rendimiento de penicilina respecto de sustrato</i>	Y_{ps}	g-Penicilina/g-sustrato	0,4 – 1,0	0,97
<i>Velocidad máxima de producción de penicilina</i>	$\pi_{m\acute{a}x}$	h^{-1}	0,003 – 0,015	0,011
<i>Constante de saturación de producción de penicilina</i>	K_p	$g L^{-1}$	0,00001 – 0,00020	0,00010
<i>Velocidad máxima de consumo de sustrato para el mantenimiento</i>	$\zeta_{m\acute{a}x}$	h^{-1}	0,014 – 0,029	0,020
<i>Constante de saturación para el mantenimiento celular</i>	K_s	$g L^{-1}$	0,00001 – 0,0002	0,0001
<i>Velocidad de hidrólisis de penicilina</i>	K_h	h^{-1}	0,002 – 0,010	0,002

Según lo anteriormente dicho para el proceso industrial, la política de operación del proceso es establecida para maximizar la cantidad de penicilina obtenida al final del cultivo. En términos matemáticos, la función objetivo toma la siguiente expresión:

$$J = P(t_f) \square V(t_f) \quad (2.3.2)$$

donde t_f es el tiempo final del proceso.

La política de operación $\wp(w, t)$ es definida mediante el vector w cuyos componentes son de dos tipos: un subconjunto de parámetros que hacen que el perfil de alimentación se modifique durante el cultivo en función del tiempo, y otro que representa las condiciones que pueden ser modificadas en cada experimento pero que una vez fijadas son invariantes durante la corrida, tales como la concentración de sustrato en la alimentación y la frecuencia de descargas. Para el primer sub-conjunto han sido propuestas diferentes representaciones del perfil de alimentación. En el presente ejemplo son evaluadas estrategias de alimentación constante y variables descritas por polinomios inversos de 1er, 2do y 3er grado. Estas funcionalidades han sido

intencionalmente escogidas para evitar inconvenientes en los *GSA* de la política de operación debido a la correlación entre los parámetros.

Caso 1: Perfil de alimentación constante

La política de alimentación del biorreactor consiste en aplicar un caudal constante a partir del instante t_0 . En términos matemáticos se utiliza la siguiente ecuación:

$$F_{in} = A; \quad t \geq t_0 \quad (2.3.3)$$

la cual queda definida por dos parámetros: A y t_0 .

En Tabla 2.3.2 son mostrados los valores superiores e inferiores permitidos para la variación de los parámetros operativos del sistema. Además, están indicados los valores resultantes de las sucesivas políticas implementadas luego del diseño “óptimo” en cada iteración. También se muestra la producción estimada y la $std[J]$.

Por simple inspección, se considera que el método ha convergido a una política de operación en la segunda iteración ya que se cumple la condición $std[J_{experimental}] \geq std[J_{estimada}]$.

Tabla 2.3. 2: Parámetros de la política de operación para caudal de alimentación constante

Parámetro	Valor inicial	Límite inferior	Límite superior	Óptimo 1ra iteración	Óptimo 2da iteración
<i>Velocidad alimentación</i> [L h ⁻¹]	4	0	10	5,73	5,73
t_0 [h]	0	0	24	0	0.025
t_f [h]	240	200	300	240	240
<i>Conc. Sustr. Alimentac.</i> [g L ⁻¹]	240	200	300	300	300
<i>Tiempo primera descarga</i> [h]	24	24	48	24	24
<i>Volumen descarga</i> [L]	60	30	80	80	80
<i>Frecuencia descarga</i> [h]	24	24	60	24	24
<i>Volumen inicial</i> [L]	600	500	700	500	500
<i>Penicilina obtenida, J</i> (Kg)	24,38			36,85	34,68
<i>std(J)</i> (Kg)				8,89	3,89

Luego de cada implementación de una política de operación, los datos experimentales obtenidos con el muestreo óptimo fueron utilizados para la re-estimación de los parámetros más relevantes según un *GSA*. La Tabla 2.3.3 muestra los valores de los

índices de sensibilidad de efectos principales normalizados S_i^n obtenidos en las dos iteraciones para todos los parámetros del modelo. Los parámetros más relevantes, y por lo tanto re-estimados, están indicados en negritas. La Tabla 2.3.4 consigna los valores reales implementados para el biorreactor *in silico* y los valores re-estimados de los parámetros del modelo más importantes para la variabilidad de J según el GSA.

Tabla 2.3. 3: Índices de sensibilidad para los parámetros del modelo adoptando una alimentación constante

Parámetro	S_i^n (1ra iteración)	S_i^n (2da iteración)
$\pi_{m\acute{a}x}$	0,6110	0,2402
K_h	0,2456	0,0198
Y_{ps}	0,0509	0,1654
$\zeta_{m\acute{a}x}$	0,0463	0,3501
Y_{xs}	0,0214	0,1234
k_{lis}	0,0000	0,0008
$\mu_{m\acute{a}x}$	0,0000	0,0008
K_x	0,0002	0,0008
K_d	0,0246	0,0978
K_s	0,0000	0,0010
K_p	0,0000	0,0000

Tabla 2.3. 4: Valores reales y re-estimados de los parámetros más relevantes según GSA

Parámetro	biorreactor <i>in silico</i>	1ra iteración	2da iteración
$\pi_{m\acute{a}x}$	0,011	(1,04±0,10)e-2	(9,8±0,8)e-3
K_h	0,002	(2,2±0,2)e-3	-
Y_{ps}	0,97	-	0,94±0,17
$\zeta_{m\acute{a}x}$	0,02	-	(1,43±0,2)e-3
Y_{xs}	0,52	-	0,47±0,04
K_d	0,006	-	(6,10±0,12)e-3

Las estrategias de muestreo optimizadas fueron:

$$1^{\circ} \text{ iter} \begin{cases} \psi^{opt}(h) : 17, 34, 141, 157, 173, 189, 205, 221, 233, 239 \\ M = 2.57 \square 10^{-38} \end{cases}$$

$$2^{\circ} \text{ iter} \begin{cases} \psi^{opt}(h) = 21, 27, 33, 143, 160, 176, 193, 210, 228, 240 \\ M = 3.40 \square 10^{-39} \end{cases}$$

Caso 2: Perfil de alimentación siguiendo un polinomio inverso de 1er orden

En este caso ha sido propuesta para la política de de alimentación una función del tipo:

$$F_{in} = \begin{cases} 0, & t < t_0 \\ A \frac{t}{B+t}, & t \geq t_0 \end{cases} \quad (2.3.4)$$

la cual queda definida por 3 parámetros: A , B y t_0 .

La Tabla 2.3.5 contiene los valores superiores e inferiores permitidos para la variación de los parámetros operativos del sistema, los resultantes de las sucesivas políticas implementadas luego del diseño “óptimo”, las producciones estimadas y los valores de $std[J]$ estimados mediante *GSA*. Nuevamente se asume la convergencia del método debido a que se cumple la condición $std[J_{experimental}] \geq std[J_{estimada}]$. Sin embargo, puede ser observado el incremento significativo en la producción de penicilina-G respecto al valor de 24 Kg reportado en el trabajo de referencia (Menezes y col., 1994).

Tabla 2.3. 5: Parámetros de la política de operación para caudal de alimentación siguiendo un polinomio lineal inverso

Parámetro	Valor inicial	Límite inferior	Límite superior	Óptimo 1ra iteración	Óptimo 2da iteración
A [$L h^{-1}$]	6	3	10	4,88	4,03
B [h]	20	0	200	13,88	1,71
t_0 [h]	0	0	24	0	0
t_f [h]	240	200	300	240	240
Conc. Sustr. Alimentac. [$g L^{-1}$]	240	200	300	300	300
Tiempo primera descarga [h]	24	24	48	25	48
Volumen descarga [L]	60	30	80	80	80
Frecuencia descarga [h]	24	24	60	53,8	60
Volumen inicial [L]	600	500	700	500	500
Penicilina obtenida, J (Kg)	27,76			34,00	34,31
$std(J)$ (Kg)				5,53	2,12

La Tabla 2.3.6 consigna los valores de los índices de sensibilidad de efectos principales normalizados Si^n obtenidos en las dos iteraciones para todos los parámetros del modelo. Los parámetros más sensibles, y por lo tanto re-estimados, están indicados en negritas. La Tabla 2.3.7 muestra los valores reales implementados para el biorreactor *in*

silico y los valores re-estimados de los parámetros del modelo más sensibles en cada iteración.

Tabla 2.3. 6: Índices de sensibilidad normalizados para los parámetros del modelo adoptando una alimentación siguiendo un polinomio lineal inverso

Parámetro	Si^n (1ra iteración)	Si^n (2da iteración)
$\pi_{m\acute{a}x}$	0,7410	0,0294
K_h	0,1062	0,0009
Y_{ps}	0,0567	0,3916
$\zeta_{m\acute{a}x}$	0,0201	0,2104
Y_{xs}	0,0079	0,0175
k_{lis}	0,0035	0,0013
$\mu_{m\acute{a}x}$	0,0030	0,0042
K_x	0,0030	0,0000
K_d	0,0030	0,1225
K_s	0,0005	0,0805
K_p	0,0000	0,1418

Tabla 2.3. 7: Valores reales y re-estimados de los parámetros más relevantes según GSA

Parámetro	biorreactor <i>in silico</i>	1ra iteración	2da iteración
$\pi_{m\acute{a}x}$	0,011	(11,05±0,18)e-3	-
K_h	0,002	(2,0±0,4)e-3	-
Y_{ps}	0,97	-	0,8±0,3
$\zeta_{m\acute{a}x}$	0,02	-	0,018±0,005
Y_{xs}	0,0001	-	(1,4±0,8)e-4
K_d	0,006	-	(5,9±0,7)e-3

Las estrategias de muestreo optimizadas fueron:

$$1^{\circ} \text{ iter } \begin{cases} \psi^{opt}(h) : 5, 14, 19, 36, 183, 187, 192, 198, 202, 232 \\ M = 3.83 \square 10^{-34} \end{cases}$$

$$2^{\circ} \text{ iter } \begin{cases} \psi^{opt}(h) : 20, 121, 129, 146, 160, 178, 184, 205, 213, 221 \\ M = 3.3 \square 10^{-36} \end{cases}$$

La Figura 2.3.1 ilustra los perfiles de alimentación implementados en cada iteración junto con la política de operación utilizada por Menezes en su contribución.

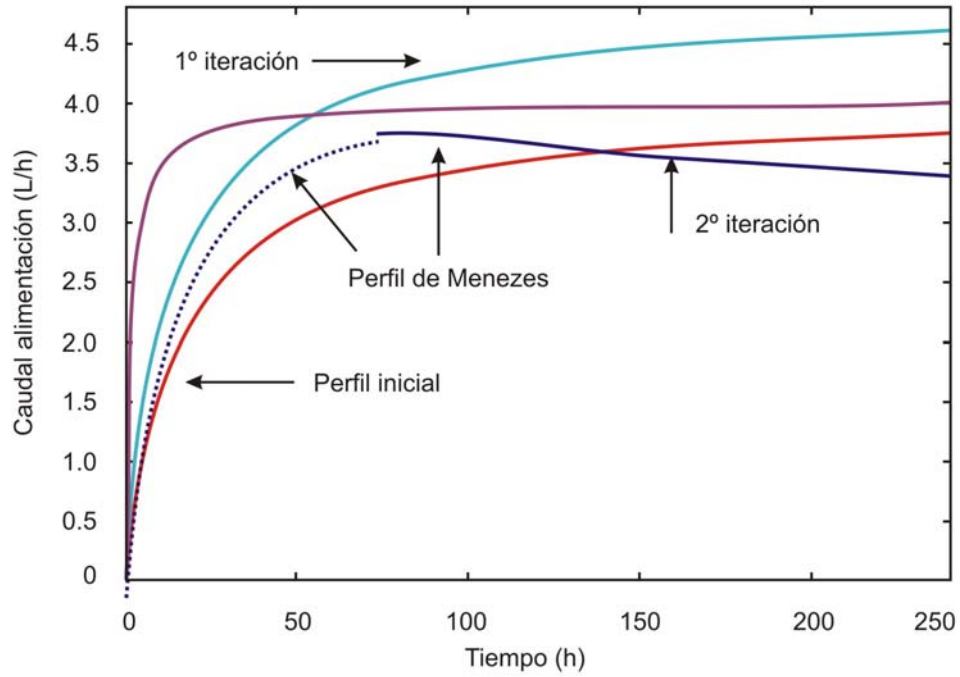


Figura 2.3.1: Políticas de alimentación implementada por Menezes y las calculadas con la metodología propuesta para la condición inicial y la 1ª y 2ª iteración asumiendo un perfil de polinomio inverso lineal.

Las Figuras 2.3.2-4 muestran los datos experimentales de las variables de estado y las trayectorias estimadas en las diferentes iteraciones.

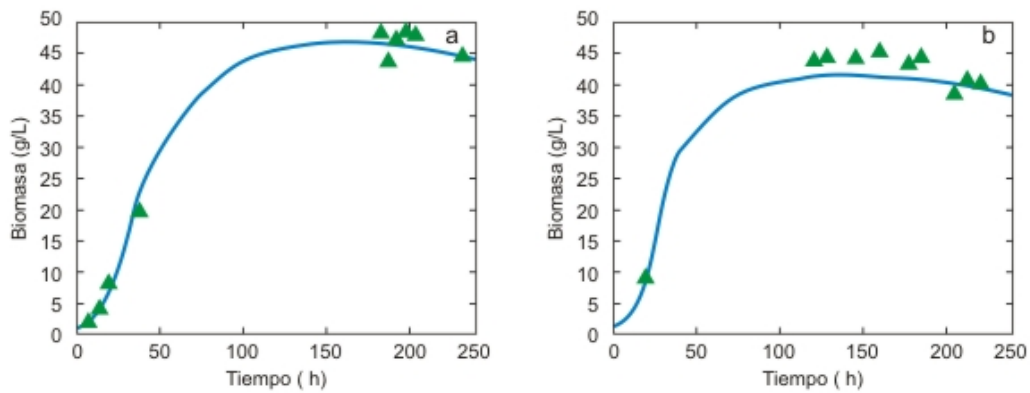


Figura 2.3.2: Datos experimentales y predicciones del modelo en la 1ª (a) y 2ª (b) iteración para la biomasa.

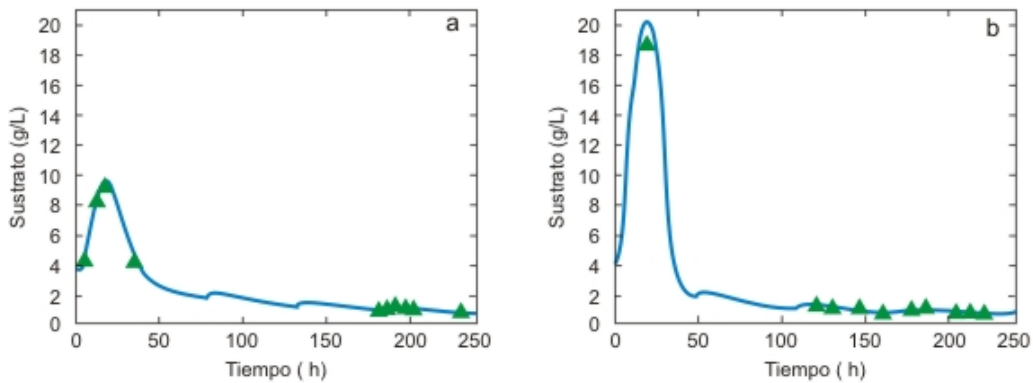


Figura 2.3.3: Tiempos de muestreo y predicciones del modelo en la 1ª y 2ª iteración para el sustrato.

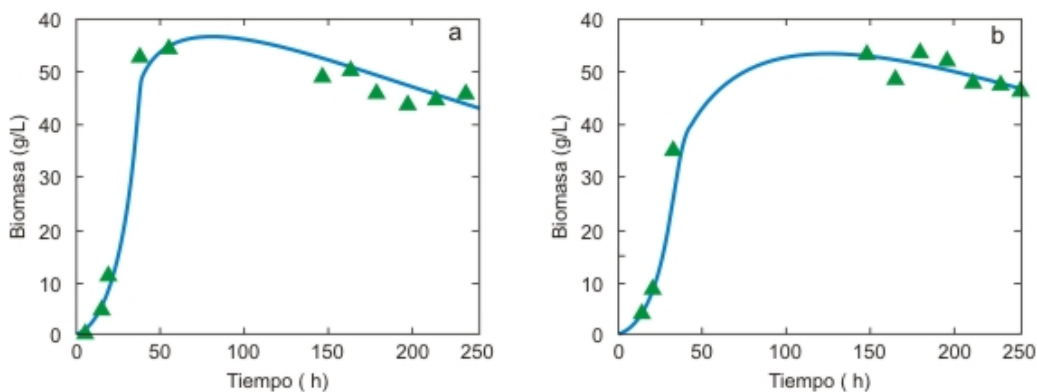


Figura 2.3.4: Tiempos de muestreo y predicciones del modelo en la 1ª y 2ª iteración para la penicilina.

Caso 3: Perfil de alimentación siguiendo un polinomio inverso de 2do orden

Con la intención de mejorar la producción de penicilina en el biorreactor se propone una alimentación siguiendo un polinomio inverso cuadrático:

$$F_{in} = \begin{cases} 0, & t < t_0 \\ \frac{At}{1 + Bt + Ct^2}, & t \geq t_0 \end{cases} \quad (2.3.4)$$

la cual queda definida por 4 parámetros: A , B , C y t_0 .

La Tabla 2.3.8 contiene los valores de los límites inferior y superior para los parámetros de la política de operación del biorreactor junto con las condiciones experimentales diseñadas utilizando la metodología de optimización iterativa propuesta. Nuevamente se

asume la convergencia del método debido a que se cumple la condición $std[J_{experimental}] \geq std[J_{estimada}]$. La producción de penicilina-G alcanza 41 Kg, representando un aumento significativo relativo a los 24 Kg reportados en el trabajo de referencia (Menezes y col., 1994).

La Tabla 2.3.9 consigna los valores de los índices de sensibilidad de efectos principales normalizados Si^n obtenidos en las dos iteraciones para todos los parámetros del modelo. Los parámetros más sensibles, y por lo tanto re-estimados, están indicados en negritas.

La Tabla 2.3.10 muestra los valores reales implementados para el biorreactor *in silico* y los valores re-estimados de los parámetros del modelo más sensibles en cada iteración.

Las estrategias de muestreo optimizadas fueron:

$$1^{\circ} \text{ iter} \begin{cases} \psi^{opt}(h) : 15, 21, 39, 55, 147, 164, 181, 198, 215, 233 \\ M = 8.06 \times 10^{-42} \end{cases}$$

$$2^{\circ} \text{ iter} \begin{cases} \psi^{opt}(h) : 20, 121, 129, 146, 160, 178, 184, 205, 213, 221 \\ M = 3.31 \times 10^{-36} \end{cases}$$

Tabla 2.3. 8: Parámetros de la política de operación para caudal de alimentación siguiendo un polinomio inverso cuadrático

Parámetro	Valor inicial	Límite inferior	Límite superior	Óptimo 1° iteración	Óptimo 2° iteración	Óptimo "real"
$A [L h^{-1}]$	0,6882	0	4,13	2,0649	2	2,0287
$B [h]$	0,1431	0,1	0,86	0,2061	0,2627	0,2000
$C [h^{-2}]$	0,0002	-0,0008	0,0012	0,0012	0,0006	0,0012
$t_0 [h]$	0	0	24	0	0	0
$t_f [h]$	240	200	300	240	240	240
Conc. Sustr. Alimentac. $[g L^{-1}]$	240	200	300	300	300	300
Tiempo primera descarga $[h]$	24	24	48	24	24	24
Volumen descarga $[L]$	60	30	80	80	80	80
Frecuencia descarga $[h]$	24	24	60	24	24	24
Volumen inicial $[L]$	600	500	700	500	500	500
Penicilina obtenida, J (Kg)	28,04			41,84	40,05	41,27
$std(J)$ (Kg)				2,0649	2	2,02

Tabla 2.3. 9: Índices de sensibilidad normalizados para los parámetros del modelo adoptando una alimentación siguiendo un polinomio inverso cuadrático.

Parámetro	Si^n (1ra iteración)	Si^n (2da iteración)
$\pi_{m\acute{a}x}$	0,7206	0,1331
K_h	0,1812	0,3114
Y_{ps}	0,0216	0,1749
$\zeta_{m\acute{a}x}$	0,0325	0,1588
Y_{xs}	0,0167	0,1111
k_{lis}	0,0000	0,0002
$\mu_{m\acute{a}x}$	0,0000	0,0002
K_x	0,0000	0,0002
K_d	0,0272	0,1099
K_s	0,0000	0,0000
K_p	0,0001	0,0002

Tabla 2.3. 10: Valores reales y re-estimados de los parámetros más relevantes según GSA

Parámetro	biorreactor <i>in silico</i>	1ra iteración	2da iteración
$\pi_{m\acute{a}x}$	0,011	(13±2)e-3	(13±2)e-3
K_h	0,002	(4±2)e-3	(4±2)e-3
Y_{ps}	0,97	-	0,97±0,1
$\zeta_{m\acute{a}x}$	0,02	-	0,014±0,010
Y_{xs}	0,52	-	0,42±0,10
K_d	0,006	-	(5±3)e-3

La Figura 2.3.5 muestra los perfiles de la política de alimentación para los diferentes experimentos. La posibilidad de presentar un máximo caudal de alimentación en los primeros momentos del cultivo, que corresponde a la fase de crecimiento exponencial de la biomasa, posibilita la obtención de una mayor cantidad de penicilina-G (aprox. 70%) que la reportada por Menezes, e incluso que la obtenible implementando un polinomio inverso lineal (Caso 2).

Puede ser observado que la metodología encuentra los valores óptimos de todos los parámetros invariantes de la de operación luego de la primera mejora del modelo (1ra iteración). Si bien en la 2º iteración el modelo predice una producción levemente menor que en la anterior iteración, debe tenerse en cuenta que la incertidumbre de ese valor es mayor e incluye a la producción del paso posterior.

En la Tabla 2.3.8 puede observarse que el método ha convergido a una política de operación “óptima” pues, debido a la incertidumbre experimental (calculada desde “réplicas” en las que ha sido introducido el error experimental en los valores de las variables de estado), la producción obtenida es la máxima del proceso. Los valores de los parámetros que definen el polinomio, sin embargo, si bien son cercanos no coinciden exactamente.

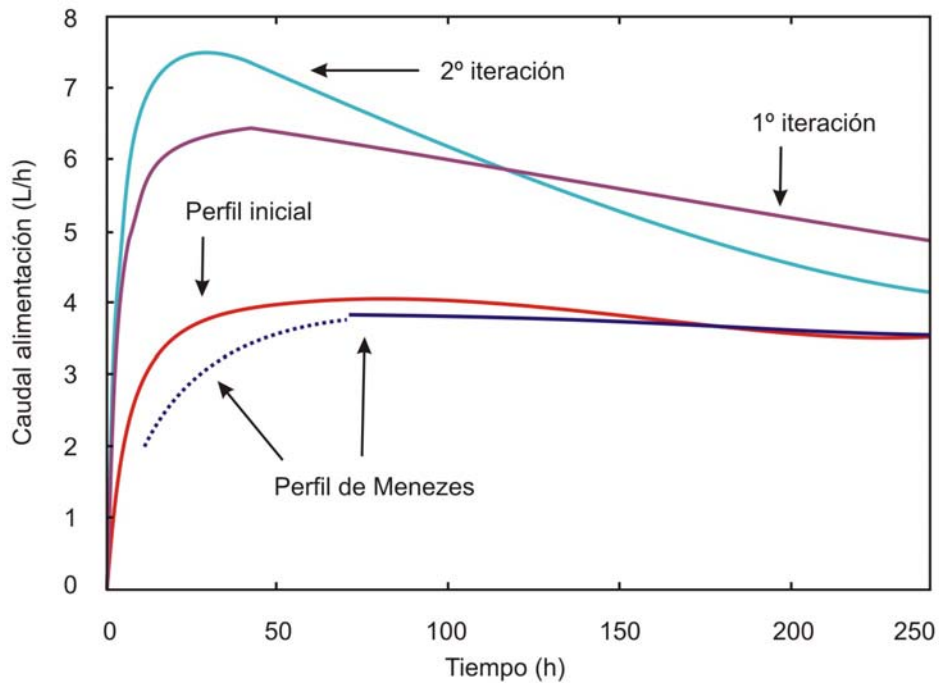


Figura 2.3.5: Políticas de alimentación implementada por Menezes y las calculadas con la metodología propuesta para la condición inicial y la 1ª y 2ª iteración asumiendo un perfil de polinomio inverso cuadrático.

Las Figuras 2.3.6-8 muestran los datos experimentales de las variables de estado y las trayectorias estimadas en las diferentes iteraciones.

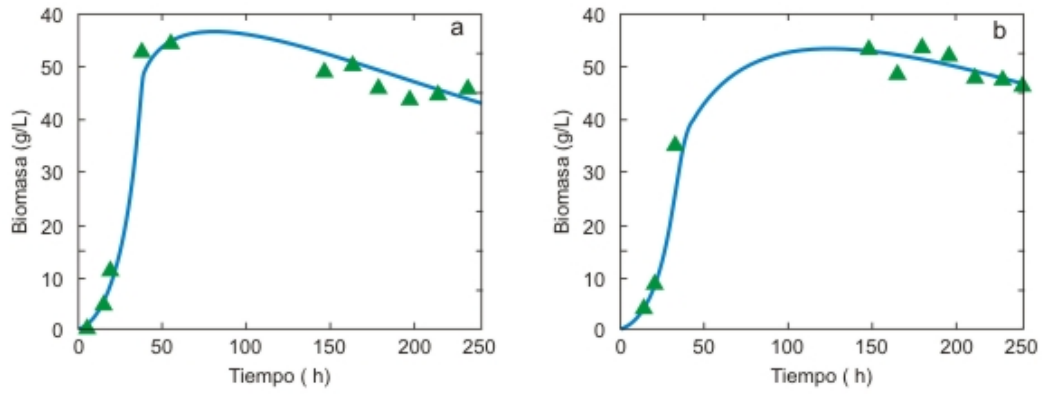


Figura 2.3.6: Datos experimentales y predicciones del modelo en la 1ª y 2ª iteración para la biomasa.

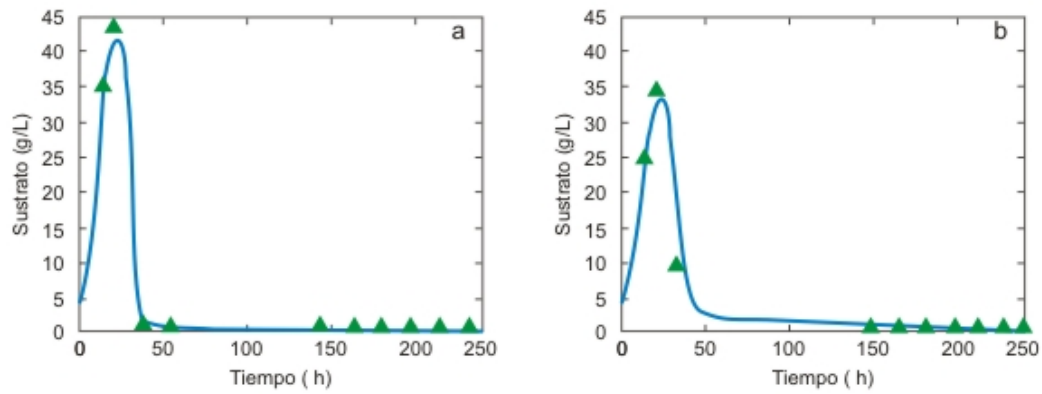


Figura 2.3.7: Datos experimentales y predicciones del modelo en la 1ª y 2ª iteración para el sustrato.

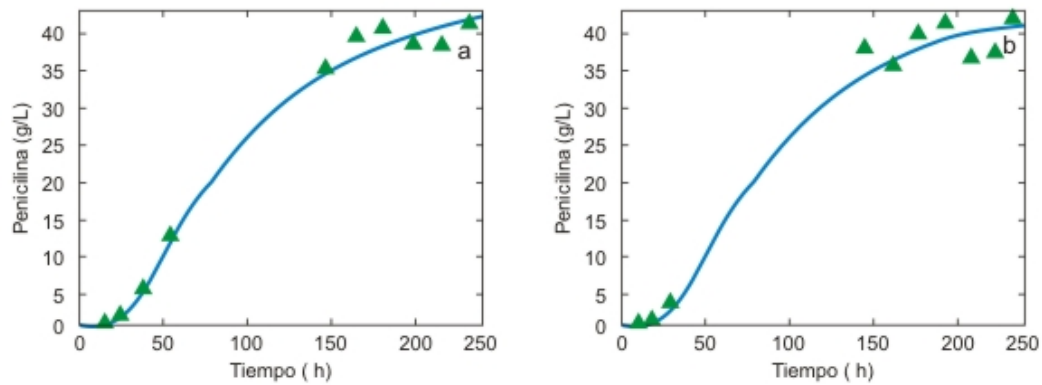


Figura 2.3.8: Datos experimentales y predicciones del modelo en la 1ª y 2ª iteración para la penicilina.

Caso 4: Perfil de alimentación del biorreactor siguiendo un polinomio inverso de 3er orden

Visto la mejora significativa de la producción de penicilina-G lograda anteriormente a partir del aumento del orden del polinomio con el que es modelado el perfil de alimentación, se decidió complejizar aún más la política de operación proponiendo la siguiente expresión para el caudal de alimentación:

$$F_{in} = \begin{cases} 0, t < t_0 \\ \frac{At}{1 + Bt + Ct^2 + Dt^3}, t \geq t_0 \end{cases} \quad (2.3.5)$$

la cual queda definida por 5 parámetros: A , B , C , D y t_0 . Es necesario destacar que no se permitió que el valor del parámetro D sea igual a cero para evitar caer en el Caso 3.

La Tabla 2.3.11 contiene los valores de los límites inferior y superior para los parámetros de la política de operación del biorreactor junto con las condiciones experimentales diseñadas utilizando la metodología de optimización iterativa propuesta. Nuevamente se asume la convergencia del método debido a que se cumple la condición $std[J_{experimental}] \geq std[J_{estimada}]$. La producción de penicilina-G alcanza aproximadamente 38 Kg, valor que determina un aumento significativo relativo a los 24 Kg reportados por Menezes y col., pero resultó ser similar al valor obtenido el Caso 3.

La Tabla 2.3.12 consigna los valores de los índices de sensibilidad de efectos principales normalizados Si^n obtenidos en las dos iteraciones para todos los parámetros del modelo. Los parámetros más relevantes, y por lo tanto re-estimados, están indicados en negritas.

La Tabla 2.3.13 muestra los valores reales implementados para el biorreactor *in silico* y los valores re-estimados de los parámetros del modelo más sensibles en cada iteración. Como puede observarse, los parámetros Y_{ps} y K_d finales no convergieron a los valores implementados en el modelo *in silico* lo que podría explicar la diferencia en la función objetivo alcanzada en la 2º iteración y la real (Tabla 2.3.11).

Las estrategias de muestreo optimizadas fueron:

$$1^{\circ} \text{ iter } \begin{cases} \psi^{opt}(h): 9, 23, 29, 35, 41, 47, 53, 66, 159, 186, 202, 218 \\ M = 1.40 \times 10^{-45} \end{cases}$$

$$2^{\circ} \text{ iter} \begin{cases} \psi^{opt}(h) : 11, 21, 27, 39, 55, 72, 82, 172, 179, 185, 191, 197 \\ M = 6.48 * 10^{-42} \end{cases}$$

Tabla 2.3. 11: Parámetros de la política de operación para caudal de alimentación siguiendo un polinomio inverso de 3° grado.

Parámetro	Valor inicial	Límite inferior	Límite superior	Óptimo 1° iteración	Óptimo 2° iteración	Óptimo "real"
A [L h ⁻¹]	0.2048	0.2151	0.3648	0.3648	0.3647	0.3648
B [h]	-0.0451	-0.0682	-0.0373	-0.0458	-0.0458	-0.0458
C [h ⁻²]	$1.4 \cdot 10^{-3}$	$1.8 \cdot 10^{-3}$	$2.7 \cdot 10^{-3}$	$1.8 \cdot 10^{-3}$	$1.8 \cdot 10^{-3}$	$1.8 \cdot 10^{-3}$
D [h ⁻³]	$-4.4 \cdot 10^{-6}$	$-6.0 \cdot 10^{-6}$	$-1.6 \cdot 10^{-6}$	$-5.7 \cdot 10^{-6}$	$-5.6 \cdot 10^{-6}$	$-5.7 \cdot 10^{-6}$
t_0 [h]	0	0	24	1.72	2.7	0.1464
t_f [h]	240	200	300	240	240	240
Conc. Sustr. Alimentac. [g L ⁻¹]	240	200	300	300	300	300
Tiempo primera descarga [h]	24	24	48	24	24	24
Volumen descarga [L]	60	30	80	54.61	54	54.74
Frecuencia descarga [h]	24	24	60	24	24	24
Volumen inicial [L]	600	500	700	500	500	500
Penicilina obtenida, J (Kg)	34.89			38.68	35.60	37.67
$std(J)$ (Kg)				5.13	1.72	1.87

Tabla 2.3. 12: Índices de sensibilidad normalizados para los parámetros del modelo adoptando una alimentación siguiendo un polinomio inverso de 3° grado

Parámetro	Si^n (1ra iteración)	Si^n (2da iteración)
$\pi_{m\acute{a}x}$	0,7166	0,2194
K_h	0,2133	0,2972
Y_{ps}	0,0182	0,1732
$\zeta_{m\acute{a}x}$	0,0360	0,2424
Y_{xs}	0,0112	0,0124
k_{lis}	0,0001	0,0000
$\mu_{m\acute{a}x}$	0,0001	0,0001
K_x	0,0000	0,0001
K_d	0,0044	0,0550
K_s	0,0001	0,0002
K_p	0,0000	0,0000

Tabla 2.3. 13: Valores reales y re-estimados de los parámetros más relevantes según GSA

Parámetro	biorreactor <i>in silico</i>	1ra iteración	2da iteración
$\pi_{m\acute{a}x}$	0,011	$(10,4\pm 1,0)e-3$	$(10,8\pm 0,8)e-3$
K_h	0,002	$(2,2\pm 0,3)e-3$	$(2,5\pm 1,2)e-3$
Y_{ps}	0,97	-	$0,88\pm 0,17$
$\zeta_{m\acute{a}x}$	0,02	-	$0,015\pm 0,003$
Y_{xs}	0,52	-	$0,88\pm 0,17$
K_d	0,006	-	$0,015\pm 0,003$

La Figura 2.3.9 muestra que las políticas de operación diseñadas a partir del modelo convergen al mismo perfil de alimentación a pesar de ser calculadas utilizando diferentes parametrizaciones. Sin embargo, nuevamente se pone evidencia el hecho de poder estimar más exactamente el índice de performance a medida que el modelo es actualizado con datos experimentales obtenidos de la aplicación de estrategias de muestreo óptimas.

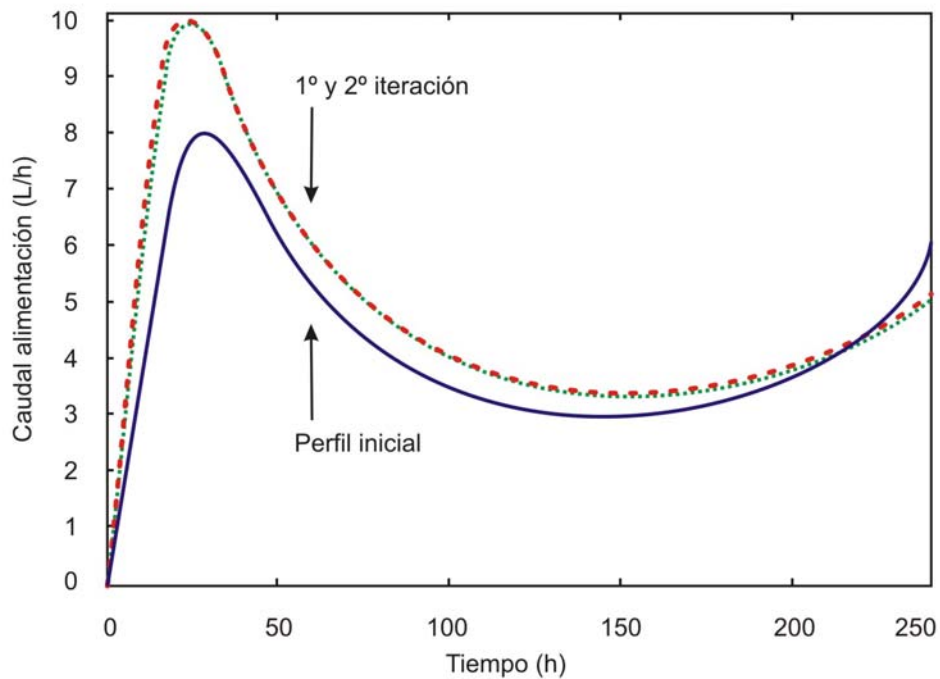


Figura 2.3.9: Políticas de alimentación calculadas con la metodología propuesta para la condición inicial y la 1ª y 2ª iteración asumiendo un perfil de polinomio inverso de grado 3.

Las Figuras 2.3.10-12 muestran los datos experimentales de las variables de estado y las trayectorias estimadas en las diferentes iteraciones.

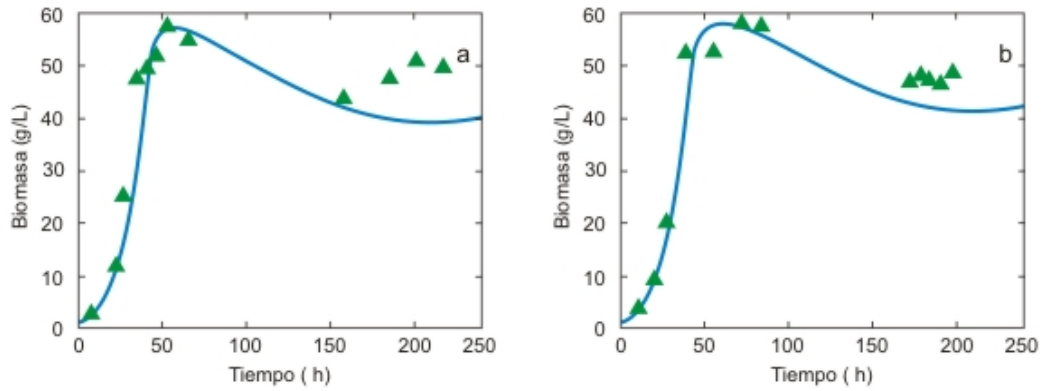


Figura 2.3.10: Datos experimentales y predicciones del modelo en la 1ra y 2da iteración para la biomasa.

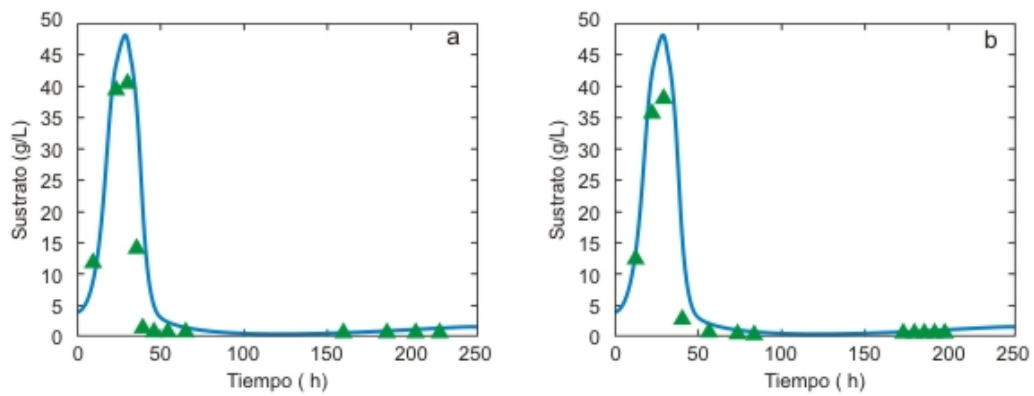


Figura 2.3.11: Datos experimentales y predicciones del modelo en la 1ra y 2da iteración para el sustrato.

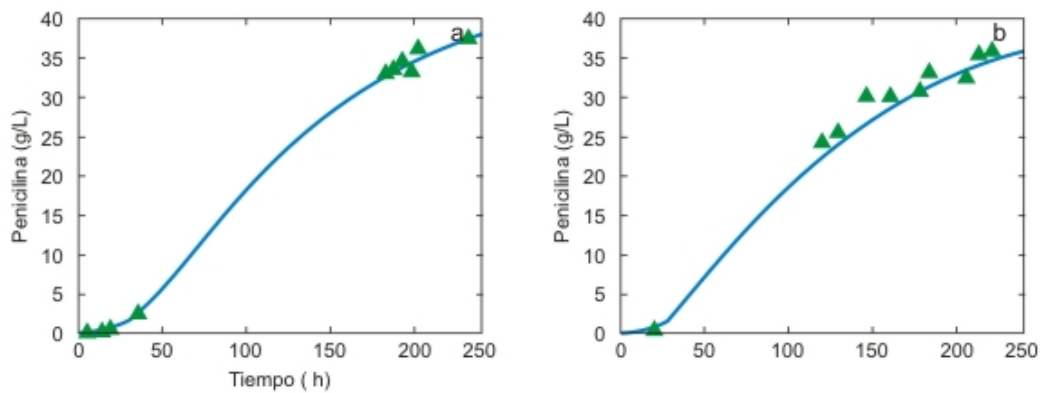


Figura 2.3.12: Tiempos de muestreo y predicciones del modelo en la 1ra y 2da iteración para la penicilina.

2.3.2 Conclusiones

La metodología de diseño de experimentos dinámicos para optimización basada en modelos de tendencias ha sido implementada para la optimización iterativa de la producción de penicilina en un cultivo de *P. crysogenum*.

El modelo reportado por Menezes y col. (1994) para describir la producción industrial de penicilina fue implementado como biorreactor *in silico*. Ese mismo modelo fue utilizado como herramienta para la optimización para analizar el caso de contar con un modelo estructuralmente correcto pero con una gran incertidumbre paramétrica.

La utilidad de la metodología propuesta ha sido ejemplificada desarrollando cuatro casos de estudio en los que se han adoptado diferentes formulaciones para definir el perfil de alimentación del biorreactor operado en modo *fed-batch*.

Del análisis de los resultados obtenidos puede concluirse que:

- El algoritmo ha demostrado ser útil para la mejora de la política de operación del biorreactor independientemente de la forma en que el perfil de alimentación es definido.
- La convergencia de la metodología a una política cuya producción es indistinguible de la óptima es rápida ya que sólo fue necesario el diseño de dos experimentos dinámicos.
- El método converge a políticas de operación óptimas pero el valor de la función objetivo depende del espacio W definido *a priori* por el experimentador.
- La incertidumbre con la que el índice de performance del proceso es estimada disminuye progresivamente en todos los casos a medida que el modelo es actualizado en cada iteración cumpliendo la condición descenso suficiente del algoritmo de identificación-optimización (Definición 2, ec. 2.26 y 2.27, sección 2.2.3).
- Los intervalos de variación de los parámetros re-estimados durante la optimización cumplen con condición de sumidero de Θ (Definición 1, ec. 2.25 sección 2.2.3).
- Independientemente de la forma matemática de la política de alimentación, los parámetros cinéticos más relevantes para explicar la variabilidad de la función objetivo han sido $\pi_{máx}$ y K_h , lo cual es esperable teniendo en cuenta que son los parámetros que definen la concentración de penicilina en el medio de cultivo. En la segunda iteración, debido a que la mayor parte de la variabilidad ha sido

reducida los perfiles de sensibilidad varían dependiendo de la política de alimentación. Sin embargo, $\pi_{máx}$ y K_h siguen apareciendo como relevantes (casos 3 y 4) y, dependiendo del caso analizado, son incluidos con importancias parecidas los parámetros Y_{ps} , $\zeta_{máx}$, K_p .

2.4. Ejemplo 2: Optimización iterativa de la producción de la hormona peptídica interleukina-11 recombinante en *Escherichia coli*

La expresión de proteínas recombinantes en *E. coli* está ampliamente difundida tanto para estudios a escala laboratorio como para la producción de proteínas de alto valor agregado a escala industrial (Choi y col., 2006). En general, los cultivos de esta bacteria para la producción de proteínas recombinantes son realizados en biorreactores operados en modo *fed-batch* hasta que una alta concentración celular es desarrollada y luego la expresión de la proteína heteróloga de interés es inducida por el agregado de una sustancia o el cambio en las condiciones de cultivo. Debido a eso, las condiciones del medio de cultivo deben permitir tanto un crecimiento óptimo como la síntesis del producto proteico.

Uno de los principales inconvenientes para lograr elevados rendimientos es la acumulación en el cultivo de metabolitos secundarios, principalmente acetato, que son producidos y liberados al medio de cultivo e inhiben el crecimiento celular. En el caso particular del acetato, este sub-producto es sintetizado por *E. coli* básicamente debido a dos causas: *i*- la falta de oxígeno disuelto y por lo tanto carencia de aceptores finales de electrones en los procesos respiratorios; *ii*- exceso de fuente de carbono, generalmente glucosa, lo que produce el llamado *over flow* metabólico o efecto Crabtree (Luli y Strohl, 1990) que es la imposibilidad de oxidar completamente la glucosa.

En ese contexto, el contar con una metodología para el diseño de experimentos dinámicos para la optimización de la producción de proteínas de alto valor agregado se vuelve una herramienta útil para abordar un rápido escalado y puesta en marcha de bioprocesos novedosos. Además, debido a la complejidad de los procesos metabólicos relatados sintéticamente, a la incertidumbre y poca reproducibilidad de los cultivos celulares en general, el desarrollo de un modelo detallado que incorpore tales fenómenos puede ser económicamente costoso y prácticamente muy difícil.

En el presente ejemplo, el Diseño de Experimentos basado en Modelos para Optimización es aplicado para la mejora en la producción de una hormona peptídica producida en *E. coli*. Para ello un modelo estructurado tetracompartimentalizado (Tang y col., 2007) ha sido implementado como biorreactor *in silico* y una biblioteca de cuatro modelos de tendencias (no estructurados) ha sido propuesta como “caja de herramientas” para la optimización iterativa del proceso. En cada iteración, los modelos

son mejorados y luego es seleccionado el mejor para ser utilizado en el diseño del próximo experimento.

Lo que se intenta estudiar es la posibilidad de mejora efectiva de la política de operación de un bioprocesos cuando están presentes tanto la incertidumbre estructural como la paramétrica.

Utilización de una biblioteca de modelos de tendencia para optimización

La metodología será aplicada para la optimización de la producción de rIL-11 utilizando *E. coli* como sistema de expresión. Como biorreactor *in silico* ha sido implementado un modelo estructurado tetracompartimentarizado (Tang y col. 2007) que describe las trayectorias de las concentraciones de las variables de estado: biomasa X , sustrato S , proteína recombinante intracelular P .

$$\frac{dV}{dt} = F_{in}; \quad D = \frac{F_{in}}{V}$$

$$\frac{dX}{dt} = \mu X - DX; \tag{2.4.1}$$

$$\frac{dS}{dt} = -\frac{\mu}{Y_{xs}} X - f(X, t) + D(S_{feed} - S);$$

$$\frac{dP}{dt} = r_p - \mu P$$

$$\text{First order: } \mu = \mu_{\max} S; \quad f(X, t) = 0 \tag{2.4.2}$$

$$\text{Monod: } \mu = \mu_{\max} \frac{S}{K_s + S}; \quad f(X, t) = 0 \tag{2.4.3}$$

$$\text{Contois: } \mu = \mu_{\max} \frac{S}{K_x X + S}; \quad f(X, t) = 0 \tag{2.4.4}$$

$$\text{Maintenance: } \mu = \mu_{\max} \frac{S}{K_s + S}; \quad f(X, t) = mX \tag{2.4.5}$$

La biblioteca de modelos de tendencia para optimización está compuesta por cuatro modelos no estructurados que difieren en la forma de describir matemáticamente la cinética de crecimiento celular (ec. 2.4.1-5). Esa biblioteca será utilizada para seleccionar el mejor modelo para el diseño de experimentos dinámicos para optimización. En cada iteración, todos los modelos son mejorados mediante la re-

estimación de sus parámetros más relevantes respecto a la estimación de la función objetivo J del bioproceso evaluados a través de una FP utilizando GSA. Dicha re-estimación es realizada utilizando datos experimentales obtenidos de aplicar una estrategia de muestreo óptima.

El índice de *performance* del proceso ha sido definido como la concentración de rIL-11 en el medio de cultivo:

$$J(t) = P(t) \square X(t) \quad (2.4.6)$$

La política de alimentación del biorreactor propuesta ha sido la de un perfil que se corresponde con un polinomio inverso de grado 2:

$$F_{in} = \begin{cases} 0 & t < t_0 \\ \frac{At}{1+Bt+Ct^2} & t \geq t_0 \end{cases} \quad (2.4.7)$$

La incertidumbre estructural de la biblioteca de modelos respecto al biorreactor *in silico* es total debido a la diferente concepción en la representación de la biomasa y es medida para cada modelo según:

$$\varepsilon_{\langle j \rangle \rightarrow J} = \frac{1}{n_{sp}} \left[\left(\langle \hat{J}_i \rangle - J_i \right)^T W_{ii} \left(\langle \hat{J}_i \rangle - J_i \right) \right]; \quad i = 1, 2, \dots, n_{sp} \quad (2.4.8)$$

La incertidumbre paramétrica inicial de los modelos ha sido definida como $\pm 50\%$ del valor nominal de cada parámetro del modelo luego de una estimación utilizando los datos experimentales publicados por Tang y col. (2007) y asumiendo una distribución uniforme en todo el intervalo de variación. La amplitud inicial de esta incertidumbre debe ser suficientemente amplia como para no acotar las posibilidades de exploración de los modelos pero teniendo en cuenta de no tornar impráctica la implementación de la metodología. Los valores nominales de los parámetros de los modelos de la biblioteca de modelos para optimización utilizados para la primera iteración son mostrados en la tabla 2.4.1.

Como primer experimento dinámico una política de operación conservadora ha sido propuesta definiendo un caudal de alimentación del biorreactor constante y una estrategia de muestreo equidistanciada.

Las condiciones de operación óptimas del biorreactor para las diferentes iteraciones diseñadas utilizando la metodología propuesta son mostradas en la tabla 2.4.2. En esa tabla también puede ser observada la mejora significativa de la producción de rIL-11 luego de seis iteraciones del algoritmo, el cual se considera que ha convergido debido a

que se ha cumplido la condición: $J_{experimental}^{n-1} = J_{experimental}^n$. Además, también puede observarse la casi convergencia en la política: $\phi^{n-1}(w^{n-1}, t) \approx \phi^n(w^n, t)$.

Tabla 2.4. 1: Intervalos de incertidumbre del modelo de Menezes (Menezes, 1994) y arametrización del biorreactor *in silico*

Parameter	Unit	Model			
		1 st Order	Monod	Contois	Maintenance
μ_{max}	h^{-1}	0.2000	0.6301	0.5607	0.5261
K_s	$g L^{-1}$	2.0184	1.4956	-	0.7190
Y_{xs}	$g_{biomass} g_{substrate}^{-1}$	0.3982	0.4506	0.4826	0.4464
K_p^{max}	$g L^{-1}$	0.0759	0.0629	0.0557	0.0536
KI_p	$g (L h)^{-1}$	0.0877	0.0609	0.0627	0.0600
K_x	$g_{substrate} g_{biomass}^{-1}$	-	-	1.7291	-
m	h^{-1}	-	-	-	0.0100

Tabla 2.4. 2: Parámetros de la política de operación para caudal de alimentación constante

Parameter	Units	Initial Condition	1 st iter	2 nd iter	3 rd iter	4 th iter	5 th iter	6 th iter
Z	$L h^{-1}$	1	-	-	-	-	-	-
A	$L h^{-2}$	0	0.0544	0.0121	0.0911	0.1431	0.2385	0.2389
B	h^{-1}	-	$3 \cdot 10^{-4}$	$3 \cdot 10^{-4}$	$3 \cdot 10^{-4}$	$3 \cdot 10^{-4}$	$3 \cdot 10^{-4}$	$3 \cdot 10^{-4}$
C	h^{-2}	-	$3 \cdot 10^{-4}$	$4 \cdot 10^{-4}$	$8 \cdot 10^{-4}$	0.0156	0.0222	0.0223
S_f	$g L^{-1}$	10	30	30	30	30	30	30
t_{feed}	h	6	0	0	3.19	4.05	5	5
t_{ind}	h	4	4	4	4	4	4	4
t_f	h	12	16	16	16	16	16	16
V_0	L	6	5.31	10	5	6.61	5	5
X_0	$g L^{-1}$	0.05	0.1	0.1	0.1	0.1	0.1	0.1
S_0	$g L^{-1}$	6	3.93	7	7	7	7	7
$J \cdot V_f$	g	1.60	6.06	3.75	7.15	6.40	7.22	7.22

Los perfiles de alimentación para las diferentes iteraciones son mostrados en la figura 2.4.1.

Al observar la tabla 2.4.2 puede apreciarse que la incorporación de incertidumbre estructural produce un cambio en el comportamiento del método respecto a la mejora monótona observada en el ejemplo anterior.

El algoritmo ha diseñado experimentos cuyas producciones reales no se corresponden con mejoras respecto de la iteración anterior inmediata, esto puede ser explicado por la parametrización incorrecta del modelo de tendencia utilizado para el cálculo de dicho experimento. Sin embargo, es importante notar que la exploración de esas zonas de

operación produce una mejor parametrización de los modelos *a posteriori* de la re-estimación de los parámetros más relevantes en la estimación de J .

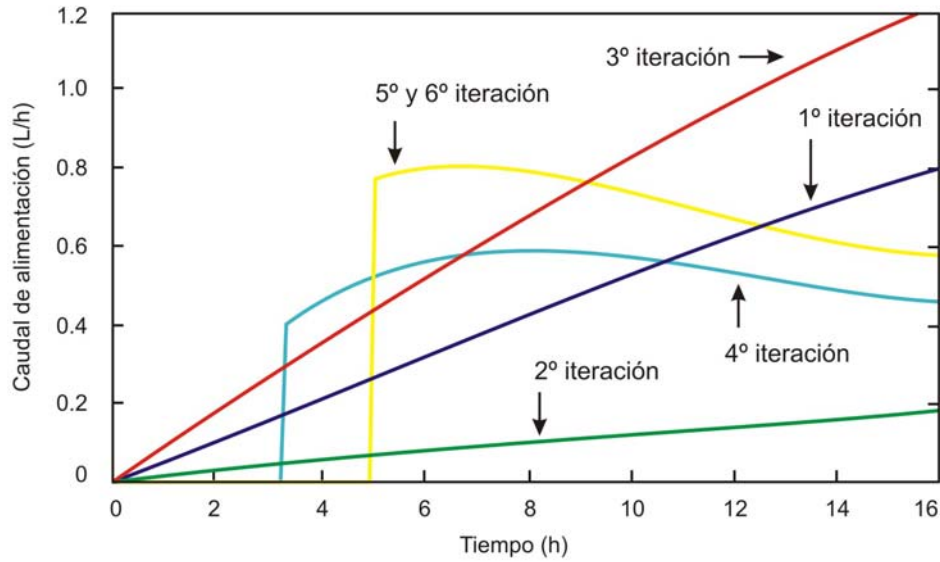


Figura 2.4.1: Políticas de alimentación implementada utilizando la metodología propuesta.

Las trayectorias experimentales y las estimadas por los diferentes modelos de tendencia mejorados luego de cada iteración son mostradas en las figuras 2.4.2-5 en las que además es listada la estrategia de muestreo óptima para cada caso.

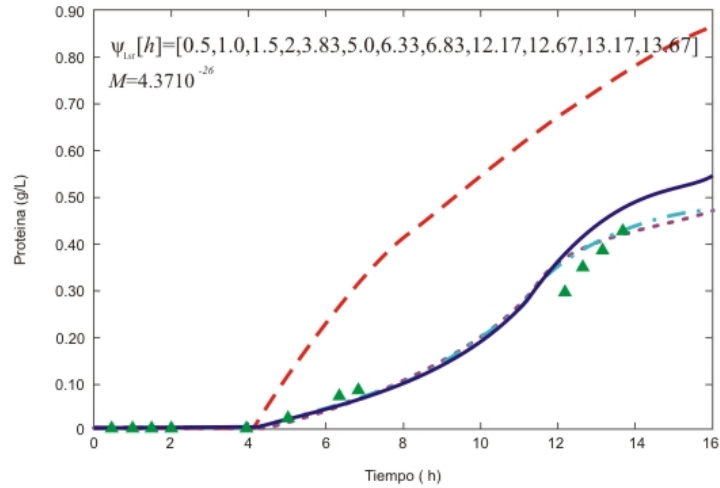


Figura 2.4.2: Trayectorias predichas por los modelos 1st order (- -), Monod (-), Contois (-.) and Maintenance (-.) y “datos experimentales” del biorreactor *in silico* (Δ) para la iteración 1.

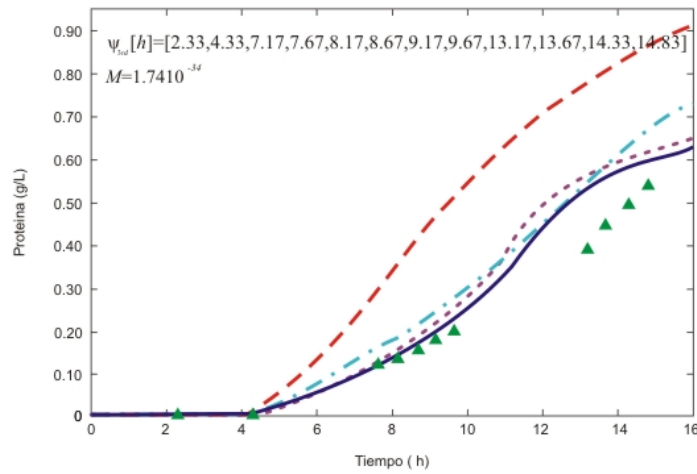


Figura 2.4.3: Trayectorias predichas por los modelos 1st order (- -), Monod (-), Contois (-.) and Maintenance (-.) y “datos experimentales” del biorreactor *in silico* (Δ) para la iteración 3.

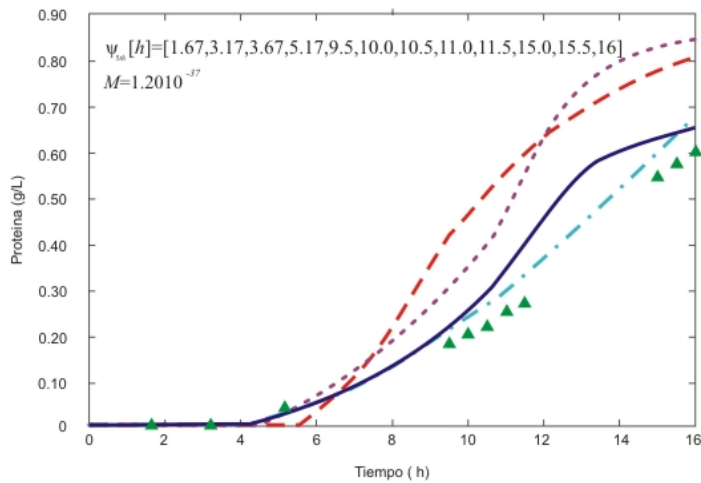


Figura 2.4.4: Trayectorias predichas por los modelos 1st order (- -), Monod (-), Contois (-.) and Maintenance (-.) y “datos experimentales” del biorreactor *in silico* (Δ) para la iteración 5.

Mejora de la biblioteca de Modelos para Optimización

La mejora de la biblioteca de modelos para optimización ha sido realizada luego de la identificación de los parámetros que mayor variabilidad introducían en estimación de la función objetivo del bioprocesos. La priorización de parámetros fue producto de un GSA a cada modelo respecto a sus espacios de incertidumbre paramétrica Θ . Los perfiles de sensibilidad de la performance a lo largo de la optimización son mostrados en las tablas 2.4.3 y 2.4.4 a partir de utilizar los índices de sensibilidad normalizados S_i^n .

Como puede ser observado, en las primeras 2 iteraciones los modelos Monod, Contois y Maintenance coinciden en el hecho que sus parámetros más inciertos son Y_{xs} y KI_p , respectivamente. Como era esperable, estos parámetros están íntimamente relacionados con la estimación de la concentración de rIL-11 en el biorreactor.

Excepto para el modelo 1st order en la iteración 2, para todos los modelos de la biblioteca en las primeras 3 iteraciones es observado un aumento en el grado de explicación de la incertidumbre de J por parte de parámetros individuales. Sin embargo, a partir de la iteración 4 comienzan aparecer combinaciones de parámetros para ser re-estimados y mayor dispersión en cuanto a los S_i^n .

Lo último indicaría que la incertidumbre paramétrica estaría distribuida en proporciones más igualitarias o lo que es lo mismo, todos los parámetros serían conocidos con el mismo grado de precisión respecto a su repercusión en la estimación de J .

Esta perspectiva es muy diferente a la clásica, en la que la precisión de un parámetro es medida respecto a la dispersión individual. Como para ejemplificar: en la tercera iteración, para el modelo Maintenance los S_i^n de los parámetros K_s y Y_{xs} son 0.013 y 0.017, respectivamente; esto significa que la incertidumbre en el conocimiento de sus valores reales no tiene una gran repercusión sobre el índice de performance. Sin embargo, los límites del intervalo de confianza de K_s es del $\pm 50\%$ respecto a su valor medio, ya que nunca fue re-estimado, mientras que para el caso de Y_{xs} los límites del intervalo de confianza están sólo a $\pm 0.50\%$ (tabla 2.4.5).

A partir de la iteración 3 los perfiles de sensibilidad varían entre los modelos dependiendo de las especificidades de cada uno. Esto puede ser atribuido a que, una vez estimados con una exactitud aceptable los parámetros que mayor relevancia tienen en la estimación del índice de *performance*, la mejora en las estimaciones producto de una

reducción en la incertidumbre paramétrica Θ de los modelos juega un rol marginal (ver más adelante los errores). Sin embargo, existen casos de parámetros que el método considera que deberían re-estimar aunque ellos ya han sido mejorados con anterioridad; esto podría ser explicado por la existencia de una “parametrización óptima” (Srivadan, 2003) de los modelos respecto de la zona de operación en la que el proceso se desarrolla debido a que modelos y proceso son estructuralmente diferentes y por lo tanto la identificabilidad y observabilidad de los parámetros no tendrían las significaciones clásicas.

Los valores re-estimados de la biblioteca de modelos durante toda la optimización son mostrados en la tabla 2.4.5. En ella aparece más nítidamente el concepto de la “parametrización óptima” de los modelos de tendencia dependiendo de las condiciones de operación del proceso ya que en ninguno de los casos donde los parámetros han sido re-estimados más de una vez, los valores mejorados caen en los intervalos de confianza construidos *a priori*. Esto es así porque durante las re-estimaciones la propiedad de Θ de presentar un sumidero no fue impuesta, por el contrario la intención ha sido evaluar el presencia de dicha propiedad en los espacios de incertidumbre paramétrica de los modelos para optimización. Sin embargo, el mejor modelo al producirse la convergencia del método ha sido Contois que justamente es aquel cuyos parámetros no presentaron un solapamiento en las re-estimaciones.

Errores de los Modelos para Optimización

Luego de la re-estimación de los parámetros más relevantes los errores paramétrico $E_{j_i \rightarrow \{j\}}$, estructural $E_{\{j\} \rightarrow j}$ y total E_{total} han sido calculados para cada modelo de la biblioteca de modelos para optimización según las formulaciones presentadas en la sección 2.1.3.

En cada iteración aquel modelo que presentaba un menor E_{total} ha sido seleccionado para el diseño del próximo experimento óptimo. Los valores de los errores de los diferentes modelos para cada iteración son mostrados en las tablas 2.4.6 y 7.

Como primera conclusión puede decirse que el crecimiento celular no puede ser asumido como siguiendo una cinética de primer orden ya que el modelo 1st Order presentó los mayores errores de los 4 que componían la biblioteca de modelos para optimización.

Cuando los datos obtenidos a partir de la implementación de la política de operación conservadora han sido utilizados para la re-estimación de los parámetros más relevantes, el modelo Monod es el que presenta el menor E_{total} y por lo tanto es el utilizado para diseñar el primer experimento dinámico para optimización. Luego de mejorada la biblioteca de modelos, el modelo de Contois aparece como el más apropiado para el diseño del próximo experimento. Sin embargo, la política óptima predicha por dicho modelo produce una disminución en la cantidad de rIL-11. Esto evidencian que la acotada incertidumbre paramétrica medida el modelo Contois no compensó los errores estructurales (que fue el mayor respecto de Monod y Maintenance) y eso ha sido crítico al momento de utilizarlo para extrapolar a otras condiciones de operación del proceso.

Además, el modelo Contois aparece como el más útil en las últimas iteraciones (hasta para diseñar el experimento 4 el modelo Maintenance es escogido pero por poca diferencia) y lleva al método a la convergencia. Ese cambio en la utilidad podría explicarse por la mejor estimación del parámetro K_x que es el que lo distingue de los demás.

Otra característica a analizar es que la presencia de un parámetro adicional en el modelo Maintenance respecto a Monod podría explicar que, excepto para la iteración 3, el primero siempre presente un error paramétrico mayor.

Por último, parecería que cuando la incertidumbre paramétrica es muy grande, esto es al principio de la optimización, las diferencias estructurales entre los modelos y de estos respecto al bioproceso no presentan una relevancia significativa. Sin embargo, a medida que los modelos son mejorados, a partir de la reducción y sesgo sistemático de los espacios de incertidumbre paramétrica, los problemas estructurales van tomando protagonismo hasta ser los responsables del estancamiento en la mejora del proceso y por lo tanto la convergencia del método.

Tabla 2.4. 3: Perfil de los índices de sensibilidad normalizados de los parámetros durante la optimización (iter. 1-3)

Parameter	S_i^n (1° iter)				S_i^n (2° iter)				S_i^n (3° iter)			
	1st Order	Monod	Contois	Maintenance	1st Order	Monod	Contois	Maintenance	1st Order	Monod	Contois	Maintenance
μ_{\max}	0,1053	0	0	0	0,4718	0,0012	0,0050	0,0003	0,0003	0,0053	0,0008	0,0157
K_s	0,0745	0,0178	-	0,0033	0,1894	0	-	0	0,3514	0	-	0,0132
Y_{xs}	0	0,5153	0,5101	0,5567	0	0,0140	0	0,0010	0	0,0721	0	0,0174
K_p^{\max}	0,2978	0,0850	0,0284	0,0609	0,3372	0,1695	0,0289	0,1902	0,6465	0,9131	0,0242	0,9455
KI_p	0,5224	0,3818	0,2576	0,3779	0,0016	0,8153	0,5957	0,8046	0,0017	0,0095	0,0001	0
K_x	-	-	0,2039	-	-	-	0,3704	-	-	-	0,9750	-
m	-	-	-	0,0012	-	-	-	0,0039	-	-	-	0,0083

Tabla 2.4. 4: Perfil de los índices de sensibilidad normalizados de los parámetros durante la optimización (iter. 4 y 5)

Parameter	S_i^n (4° iter)				S_i^n (5° iter)			
	1st Order	Monod	Contois	Maintenance	1st Order	Monod	Contois	Maintenance
μ_{\max}	0	0,2405	0,0367	0,0137	0,1578	0,1793	0,5578	0,0297
K_s	0,9950	0,3614	-	0	0,3202	0,1429	-	0
Y_{xs}	0,0010	0,3157	0,0185	0,1217	0,1375	0	0,1518	0,2011
K_p^{\max}	0,0033	0,0824	0,9263	0,1336	0,3845	0,2092	0,1046	0,2353
KI_p	0,0006	0	0	0,3953	0	0,4686	0	0,4781
K_x	-	-	0,0186	-	-	-	0,1857	-
m	-	-	-	0,3357	-	-	-	0,0558

Tabla 2.4. 5: Valores re-estimados de los parámetros más relevantes según GSA

Parameter	1st Order	Monod	Contois	Maintenance
μ_{\max}	(373±7)e-3 (2 iter)	(488,2±1,5)e-3 (initial)	(48,5±1,1)e-2 (initial) (472±2)e-3 (5 iter)	(445,2±1,4)e-3 (initial)
K_s	(295±6)e-2 (4 iter); (370±7)e-2 (5 iter)	(1380±8)e-3 (4 iter)	-	-
Y_{xs}	(68,7±1,2)e-2 (initial)	(487±3)e-3 (1 iter); (478±3)e-3 (4 iter)	(780±9)e-3 (1 iter)	(438±2)e-3 (1 iter)
K_P^{\max}	(104±2)e-3 (3 iter); (46±3)e-3 (5 iter)	(52,8±1,1)e-3 (3 iter)	(413±8)e-4 (4 iter)	(58,3±1,1)e-3 (3 iter)
KI_P	(70,7±1,4)e-3 (1 iter)	(70,1±1,3)e-3 (2 iter); (75,2±1,5)e-3 (5 iter)	(77,7±1,6)e-3 (2 iter)	(80±2)e-3 (2 iter); (94±2)e-3 (4 iter); (119±2)e-3 (5 iter)
K_x	-	-	(1131±8)e-3 (3 iter)	-
m	-	-	-	(581±1,2)e-4 (4 iter)

2.4.2 Conclusiones:

La metodología para el Diseño de Experimentos Dinámicos basado en Modelos para Optimización ha sido implementada para la optimización de la producción de rIL-11 en un biorreactor de *E. coli* operado en modo fed-batch. Como biorreactor *in silico* ha sido implementado un modelo estructurado tetracompartimentalizado propuesto por Tang y col. (2007). Se intentó optimizar dicho biorreactor utilizando la versión completa del algoritmo propuesto que incluye una biblioteca de modelos de tendencias para optimización.

La biblioteca de modelos para optimización esta compuesta por 4 modelos no estructurados y en cada iteración uno de ellos fue escogido, previa mejora mediante la re-estimación de los parámetros más relevantes, para la estimación del índice de *performance*, en base a una medida del error total que incluye tanto la incertidumbre paramétrica como estructural.

De los resultados obtenidos puede concluirse que:

- La implementación del Diseño de Experimentos Dinámicos basados en una Biblioteca de Modelos para Optimización ha resultado de utilidad para alcanzar una mejora significativa de la producción de rIL-11 en un cultivo *in silico* de *E. coli*.
- El método propuesto converge a una política de operación mejorada. Esa convergencia no se produce en forma monótona sino que han sido exploradas

zonas de operación que producen índices de *performance* bajos. Sin embargo, el conocimiento incorporado a través de esas experiencias es sumamente enriquecedor para la mejora de la parametrización de la biblioteca de modelos de tendencias.

- Los estimadores para determinar la incertidumbre paramétrica y estructural propuestos han mostrado ser útiles para seleccionar al mejor modelo de tendencias de una biblioteca para el diseño de experimentos para optimización.
- La presencia de incertidumbre estructural ha puesto de manifiesto la dependencia de la parametrización de los modelos de tendencias respecto a las condiciones en las que el sistema es operado poniendo en duda la aplicación de los conceptos de identificabilidad y observabilidad de las técnicas de modelado clásico.
- Los perfiles de sensibilidad paramétrica de los modelos de tendencias han mostrado que existen algunos pocos parámetros que definen la capacidad predictiva de aquellos y que sus valores son críticos para una buena estimación de la función objetivo por lo que deben ser re-estimados más de una vez durante la optimización.

Tabla 2.4. 6: Errores paramétrico, estructural y total para la biblioteca de modelos (iter. 1-3)

Model	inicial			1° iter			2° iter			3° iter		
	$E_{j_i \rightarrow \langle j \rangle}$	$E_{\langle j \rangle \rightarrow J}$	E_{total}	$E_{j_i \rightarrow \langle j \rangle}$	$E_{\langle j \rangle \rightarrow J}$	E_{total}	$E_{j_i \rightarrow \langle j \rangle}$	$E_{\langle j \rangle \rightarrow J}$	E_{total}	$E_{j_i \rightarrow \langle j \rangle}$	$E_{\langle j \rangle \rightarrow J}$	E_{total}
1st Order	0,0044	0,0209	0,0254	0,0049	0,0485	0,0534	0,0023	0,0064	0,0087	$4,87 \cdot 10^{-4}$	0,0465	0,0470
Monod	$7,13 \cdot 10^{-12}$	$6,23 \cdot 10^{-5}$	$6,23 \cdot 10^{-5}$	0,0034	$3,96 \cdot 10^{-4}$	0,0038	$4,32 \cdot 10^{-4}$	$5,61 \cdot 10^{-5}$	$4,88 \cdot 10^{-4}$	$5,24 \cdot 10^{-4}$	0,0029	0,0034
Contois	$1 \cdot 10^{-4}$	$5,78 \cdot 10^{-5}$	$1,58 \cdot 10^{-4}$	0,0013	0,0023	0,0033	$2,65 \cdot 10^{-4}$	0,0011	0,0014	$1,15 \cdot 10^{-4}$	0,0020	0,0021
Maintenance	1,1403	$6,35 \cdot 10^{-5}$	1,1404	0,0060	$2,79 \cdot 10^{-4}$	0,0062	$5,50 \cdot 10^{-4}$	$1,01 \cdot 10^{-4}$	$6,51 \cdot 10^{-4}$	$5,24 \cdot 10^{-4}$	0,0117	0,0119

Tabla 2.4. 7: Errores paramétrico, estructural y total para la biblioteca de modelos (iter. 4 y 5)

Model	4° iter			5° iter		
	$E_{j_i \rightarrow \langle j \rangle}$	$E_{\langle j \rangle \rightarrow J}$	E_{total}	$E_{j_i \rightarrow \langle j \rangle}$	$E_{\langle j \rangle \rightarrow J}$	E_{total}
1st Order	$6,41 \cdot 10^{-7}$	0,0279	0,0279	$1,02 \cdot 10^{-4}$	0,0189	0,0190
Monod	$6,92 \cdot 10^{-6}$	0,0011	0,0011	$1,18 \cdot 10^{-5}$	0,0069	0,0070
Contois	$1,65 \cdot 10^{-5}$	$9,85 \cdot 10^{-4}$	0,0010	$3,59 \cdot 10^{-5}$	0,0064	0,0065
Maintenance	$2,01 \cdot 10^{-4}$	0,0045	0,0047	$9,38 \cdot 10^{-4}$	0,0218	0,0227

2.5 Bibliografía:

- Asprey S. y Macchietto, S. (2000), Statistical tools for optimal dynamic model building; *Computers and Chemical Engineering*, (24) 4-7, 1261 – 1267.
- Asprey S.; Macchietto, S. (2002). Designing robust optimal dynamic experiments. *Journal of Process Control*, **12** (4), 545-556.
- Banks H.T.; Dediu S.; Ernstberger S.; Kappel F. (2009). Generalized Sensitivities and optimal experimental design. *Journal of Inverse and III-Posed Problems*, **18**(1), 25-83.
- Bellman, R.E. (1958). *Dynamic programming*, Princeton University Press.
- Bertsekas, D. P. (1995). *Dynamic programming and optimal control*, 1st Ed., Vol 1, Ch. 3, Athena Scientific.
- Bonvin, D. (1998). Optimal operation of batch reactors: a personal view. *Journal of Process Control*, **8**(5,6), 355-368.
- Brdyś, M.; Roberts, P. (1987). Convergence and optimality of modified two-step algorithm for integrated system optimisation and parameter estimation; *International Journal of Systems Science*, 18, 1305–1322.
- Chen B. y Asprey S.; On the design of optimally informative experiments for model discrimination among dynamic crystallization process models; Proceedings Foundations of Computer-Aided Process Operations; 2003; 455 – 458.
- Choi, J.H.; Keum, K.C.; Kee, S.Y. (2006). Production of recombinant proteins by high cell density culture of Escherichia coli. Review; *Chemical Engineering Science*, **61**(3), 876-885.
- Filippi, C.; Greffe, J.L.; Bordet, J.; Villiermaux, J.; Barnay, J.L.; Bonte, P.; Georgakis; C. (1986). Tendency modeling of semibatch reactors for optimization and control; *Chemical Engineerig Science*, **41**(4), 913-920.
- Filippi, C.; Bordet, J.; Villiermaux, J.; Marchal-Brasselya, S.; Georgakis, C. (1989). Batch reactor optimisation by use of Tendency models; *Computer and Chemical Engineering*; **13**(1-2), 35-47.
- Fotopoulos, J.; Georgakis, C.; Stenger Jr., H. G. (1994). Uncertainty issues in the modelling and optimisation of batch reactors with Tendency models. *Chemical Engineerig Science*, **49**(24), 5533-5547.
- Fotopoulos, J.; Georgakis, C.; Stenger Jr., H. G. (1996). Effect of the process-model mismatch on the optimisation of the catalytic epoxidation of oleic acid using Tendency models. *Chemical Engineerig Science*, **51**(10), 1899-1908.
- Fotopoulos, J.; Georgakis, C.; Stenger Jr., H. G. (1998). Use of Tendency models and their uncertainty in the design of state estimators for batch reactors. *Chemical Engineering and Processing*. **37**(6), 545-558.
- Georgakis, G. (1995). Modern tools of process control: The case of black, gray and white models. *Entropie*; **31**(194), 34-48.
- Joshi, M.; Seidel-Morgenstern, A.; Kremling, A. (2006). Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems. *Metabolic Engineering*; **8**, 447–455.

- Kadam, J.V.; Schlegel, M.; Srinivasan, B.; Bonvin, D.; Marquardt, W. (2007). Dynamic optimization in the presence of uncertainty: from off-line nominal solution to measurement-based implementation. *Journal of Process Control*, **17**(5), 389-398.
- Luli, G.W.; Strohl, W.R. (1990). Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. *Applied and Environmental Microbiology*, **56**(4), 1004-1011.
- Martínez, E. C.; Wilson, J. A. (2003). Evolutionary optimization of batch process systems using imperfect models. *Proceeding of the Indian Natural Science Academy*, **69**, 403-428.
- Menezes, J.C.; Alves, S.S.; Lemos, J.M.; Fedo de Azevedo, S. (1994). Mathematical modelling of industrial pilot-plant penicillin-G fed-batch fermentations. *Journal of Chemical Technology and Biotechnology*, **61**(2), 123–138.
- Rodriguez-Fernandez M., Kucherenko S., Pantelides C. and Shah N. (2007). Optimal experimental design based on global sensitivity analysis, *Computer-Aided Process Engineering*. **24**, 63-68.
- Polak, E. (1997). *Optimization: algorithms and consistent approximations*, Springer: New York.
- Schenker, B.; Agarwal, M. (1995). Prediction of infrequently measurable quantities in poorly modeled processes. *Journal of Process Control*, **5**(5), 329-339.
- Sedrati, Y.; Cabassud, M.; Le Lann, M.V.; Casamatta, G. (1999). Sequential experiment design strategy for kinetic parameter estimation. *Computers and Chemical Engineering*, **23**(1), S427-S430.
- Sobol', I.M. (1974). *The Monte Carlo method*; The University of Chicago Press, Ltd., USA.
- Srinivasan, B.; Bonvin, D. (2003). Convergence analysis of iterative identification and optimization schemes. *Proceedings of the ACC*, Denver, CO, USA.
- Tang, S.; Chen, J.; Zhang, Z. (2007). Structured models for recombinant human interleukin-11 fermentation. *Biochemical Engineering Journal*, **35**, 210-217.
- Terwiesch, P.; Agarwal; M.; Rippin, D.W.T. (1994). Batch unit optimization with imperfect modeling: a survey. *Journal of Process Control*, **4**(4), 238-258.
- Tsobanakis, P. (1994). Tendency Modelling and optimisation of fed-batch fermentations. Ph. D. Thesis, Lehigh University.

yo: – y ahora...¿Para adónde arranco?
ella (indignada): – Ay tío! ...vos siempre igual...no cambias más! ...
seguime que yo te guío...

Diálogos con Victoria

Capítulo 3
Optimización Experimental utilizando
Modelos de Tendencias Probabilísticos

3.1 Hipótesis de Trabajo

En cada experiencia del proceso iterativo de optimización experimental utilizando el Diseño de Experimentos Dinámicos basados en Modelos se recolectan datos que luego son utilizados para sesgar el espacio de incertidumbre paramétrica Θ de los modelos para así direccionar la búsqueda del óptimo. Esos datos también pueden aportar información sobre la topología de Θ a partir de la estimación empírica de las funciones de densidad de probabilidad de los parámetros utilizando *bootstrapping*.

3.2. Integración de *bootstrapping* en el desarrollo de modelos de tendencia probabilístico

La metodología propuesta para el Diseño de Experimentos basado en Modelos para Optimización presentado en el capítulo anterior ha demostrado ser útil en la mejora de la producción en diferentes bioprocesos. Sin embargo, para la evaluación de la incertidumbre paramétrica ha sido adoptada una visión pesimista respecto al desconocimiento de los parámetros del modelo para optimización ya que fueron asumidas funciones de densidad de probabilidad (*fdp*) uniformes para ellos al momento de realizar los *GSA*. Además, si bien luego de la re-estimación de parámetros los intervalos de confianza han sido obtenidos utilizando la técnica de Monte Carlo Filtering y no la clásica matriz de información de Fisher (que provee de las estimaciones más bajas de las varianzas paramétricas), aquellos se asumen simétricos respecto al valor promedio. Esto último no aporta información para caracterizar al espacio de incertidumbre paramétrica.

La técnica de *bootstrap* (Anexo 3.1) ha sido integrada a la estimación de parámetros en la mejora de los modelos para optimización (Figura 3.1). Esto ha sido realizado con el objetivo de aumentar el conocimiento del dominio de incertidumbre paramétrica, en cuanto a la caracterización de la estructura de las *fdp* de los parámetros de los modelos de tendencias, mediante su estimación empírica junto con la definición no paramétrica de los intervalos de confianza.

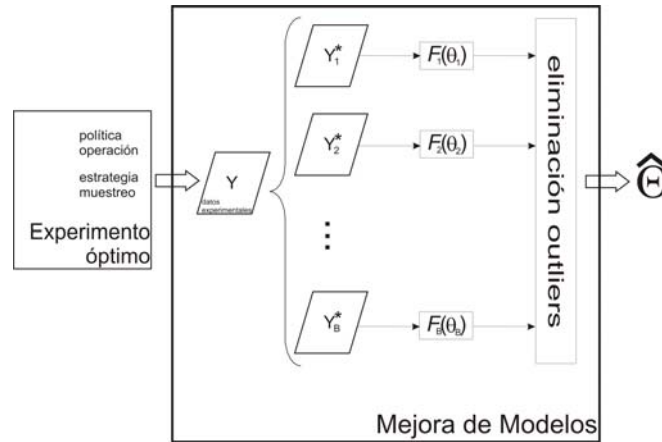


Figura 3.1: Representación pictórica de integración del bootstrapping en el Diseño de Experimentos Dinámicos para Optimización basado en Modelos Probabilísticos

3.3. Ejemplo 1: Diseño de Experimentos Dinámicos para la Optimización de la producción de penicilina en un cultivo de *Penicillium crysogenum* utilizando un modelo de tendencia probabilística. Modelo estructuralmente correcto.

Este ejemplo considera el modelo estructuralmente correcto descrito en el ejemplo 1 del capítulo 2 por sistema de ecuaciones diferenciales y condiciones dadas por las Ecs. (2.3.1). La metodología propuesta ha sido implementada de la misma forma que en aquel caso, sólo que en el presente ejemplo son introducidas las siguientes modificaciones:

- La política de alimentación del biorreactor es implementada usando la siguiente formulación: **(en la ecuación cambiar t_{feed} por t_f)**

$$F_{in} = \left\{ \begin{array}{l} 0, \quad t < t_{feed} \\ \frac{At'}{1 + Bt' + Ct'^2}, t \geq t_{feed} \\ t' = t - t_{feed} \end{array} \right\} \quad (3.3.1)$$

- El tiempo máximo de duración de la fermentación es de 300 horas.
- La dispersión de los “datos experimentales” es del 5% para cada variable de estado.
- Los índices de sensibilidad mostrados son los de efectos principales de Sobol’ S_i .

La Tabla 3.3.1 consigna los valores de los parámetros de la política de operación implementados a partir del diseño de experimentos para optimización, junto a los

valores que pertenecen al óptimo real del biorreactor. Además, son mostrados los índices de performance reales obtenidos al final de cada corrida con la incertidumbre de su estimación debido al uso de un modelo con incertidumbre paramétrica. La mejora de la *performance* del proceso es significativa respecto a la política de operación inicial ya que ha aumentado en más del 130%. A diferencia del ejemplo 1 del capítulo 2, en el presente la convergencia es alcanzada luego de cuatro experimentos para optimización. Sin embargo, hay que observar que el criterio de convergencia ha sido intensificado debido a la disminución del error “experimental” de la función objetivo.

La Tabla 3.3.2 muestra los valores de los índices de sensibilidad de efectos principales *Si* para las diferentes iteraciones. Los parámetros más relevantes, y por lo tanto re-estimados, están indicados en negritas. Puede ser observado que en las 2 primeras iteraciones existe un parámetro individual que explica más del 20% de la variabilidad del índice de *performance* mientras que luego de la estimación de los parámetros más inciertos la incertidumbre paramétrica se distribuye entre un mayor número de parámetros.

La Figura 3.3.1 muestra los perfiles de la política de alimentación para las diferentes iteraciones junto con la alimentación óptima. Se puede observar que durante el proceso de optimización el método ha calculado políticas de alimentación que fluctúan sobre y por debajo de la política de alimentación óptima y se acercan a ella. Además, según los resultados obtenidos no existen diferencias en la producción de penicilina cuando son implementadas las políticas de operación calculadas en las iteraciones 3, 4 o si la óptima es encontrada demostrando la multiplicidad de políticas que pueden llevar a una producción óptima.

Tabla 3.3. 1: Parámetros de la política de operación durante optimización experimental y los de la política de operación óptima

Parámetro	Valor inicial	Límite inferior	Límite superior	Optimo (1ra iter.)	Optimo (2da iter.)	Optimo (3ra iter.)	Optimo (4ta iter.)	Optimo “real”
A [L h ⁻²]	0,6882	0	4,13	1,5537	1,0357	1,0993	1,1153	1,2100
B [h ⁻¹]	0,1431	0,1	0,86	0,1120	0,1	0,1000	0,1006	0,1000
C [h ⁻²]	0,0002	-0,0008	0,0012	0,0012	0,0012	0,0005	0,0006	0,0007
t_{feed} [h]	0	0	24	24	0	24	24	24
t_{final} [h]	240	200	300	243	223,22	294,3	283,0	282,7
<i>Conc. Sustr. Alimentac.</i> [g L ⁻¹]	240	200	350	350	350	350	350	350
<i>Tiempo primera descarga</i> [h]	24	24	48	24	48	24	24	24
<i>Volumen descarga</i> [L]	60	30	80	80	80	80	80	80
<i>Frecuencia descarga</i> [h]	24	24	60	24	60	24	25,9	25,13
<i>Volumen inicial</i> [L]	600	500	700	500	500	500	500	500
<i>Penicilina obtenida, J</i> (Kg)	22,54			47,23	45,90	50,61	51,88	51,92
<i>std(J)</i> (Kg)				4,42	4,08	2,58	1,27	

Tabla 3.3. 2: Índices de sensibilidad para los parámetros del modelo

Parámetro	Si (1ra iter.)	Si (2da iter.)	Si (3ra iter.)	Si (4ta iter.)
$\mu_{m\acute{a}x}$	0,1818	0,0187	0,0731	0,0840
K_x	0,0282	0,0194	0	0,0134
K_d	0,0891	0	0,1051	0,1166
k_{lis}	0,0327	0,0193	0,0563	0,0845
Y_{xs}	0,1641	0,1635	0,0535	0,1024
Y_{ps}	0,0906	0,0684	0,1308	0,1325
$\pi_{m\acute{a}x}$	0,0408	0,2306	0,0450	0,0122
K_p	0,0342	0,0195	0,0708	0,0849
$\zeta_{m\acute{a}x}$	0	0,1659	0,1344	0,0705
K_s	0,0327	0,0192	0,1124	0,0847
K_h	0,3031	0,0947	0,1738	0

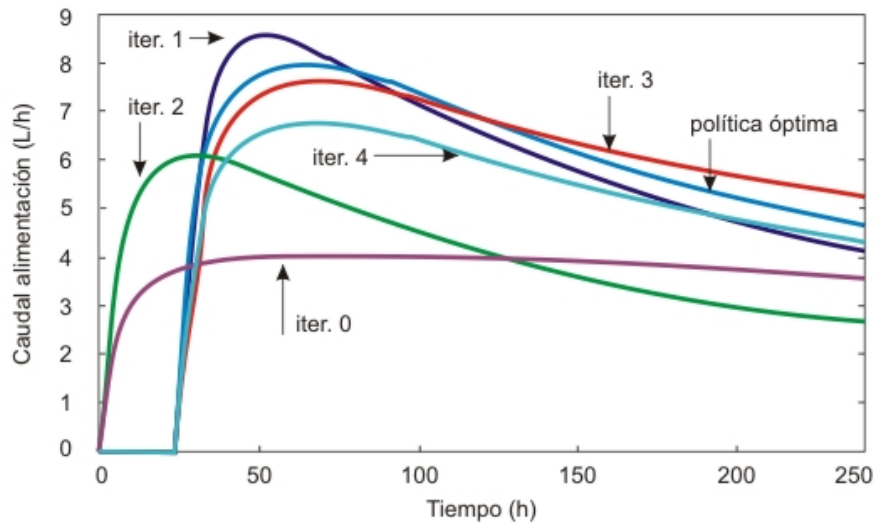


Figura 3.3.1: Políticas de alimentación implementadas para la optimización iterativa de biorreactor de producción de penicilina.

Con los datos experimentales recolectados luego de la implementación de la política de operación inicial han sido estimadas las fdp y fdp acumulada empíricas de todos los parámetros cinéticos. La Figura 3.3.2 ilustra las distribuciones que fueron construidas mediante la aplicación de la técnica de *bootstrapping* para el conocimiento de la incertidumbre paramétrica asociada al modelo para optimización. Puede observarse que utilizando relativamente poca información, esto es la obtenida de una sola experiencia, las fdp de todos los parámetros distan mucho de ser uniformes como había sido asumido en el ejemplo 1 del capítulo anterior, y que los intervalos de confianza han sido

reducidos significativamente. Las *fdp* de algunos parámetros como K_p , k_{lis} y K_s presentan una excesiva acumulación de probabilidad de ocurrencia del valor de los parámetros en un rango muy estrecho de valores, esto podría deberse a la rutina de estimación que utiliza el método Levenberg-Marquardt para optimización que fija los valores de los parámetros poco relevantes para la optimización en sus valores límites.

La Tabla 3.3.3 muestra los intervalos de confianza del 95 percentilo para los parámetros más relevantes respecto a la precisión en la estimación de la función objetivo estimados utilizando *bootstrapping*. Los intervalos finales que incluyen el valor real del parámetro considerado están resaltados. En esta nueva versión de la metodología de optimización, la sensibilidad del índice de performance ha sido calculada teniendo en cuenta las distribuciones de *fdp* empíricas de los parámetros como formadoras del espacio de incertidumbre paramétrica.

Tabla 3.3. 3: Rangos iniciales y re-estimados de los parámetros del modelo durante la optimización

Parámetro	Biorreactor "in silico"	Rango inicial	Rango re-estimado
$\mu_{m\acute{a}x}$	0,13	0,12 – 0,17	0,1251 – 0,1671 (0 iter) 0,1247 – 0,1490 (1 iter)
K_x	0,131	0,006 – 0,4	0,1224 – 0,1815 (0 iter)
K_d	0,006	(5 – 10)e-3	(5,0 – 8,1)e-3 (0 iter) (7 – 10)e-3 (3 iter) (7 – 8,2)e-3 (4 iter)
k_{lis}	0,0008	(0,01 – 8)e-3	(1,0 – 3,13)e-5 (0 iter)
Y_{xs}	0,52	0,40 – 0,58	0,4855 – 0,5799 (0 iter) 0,4508 – 0,5687 (1 iter) 0,4844 – 0,5615 (2 iter) 0,4844 – 0,5287 (4 iter)
Y_{ps}	0,97	0,4 – 1	0,5587 – 0,9772 (0 iter) 0,6531 – 0,9992 (3 iter) 0,6531 – 0,9564 (4 iter)
$\pi_{m\acute{a}x}$	0,011	(3 – 15)e-3	(1,08 – 1,28)e-2 (0 iter) (1,13 – 1,20)e-2 (2 iter)
K_p	0,0001	(1 – 20)e-5	(1 – 19,59)e-5 (0 iter)
$\zeta_{m\acute{a}x}$	0,02	0,014-0,029	(1,4 – 2,33)e-2 (0 iter) (1,4 – 1,87)e-2 (2 iter) (1,4 – 1,82)e-2 (3 iter)
K_s	0,0001	(1 – 20)e-5	(1,0 – 20)e-5 (0 iter) (1,0 – 20)e-5 (3 iter)
K_h	0,002	(2 – 10)e-3	(2,0 – 4,7)e-3 (0 iter) (2,01 – 4,43)e-3 (1 iter) (2,0 – 3,9)e-3 (3 iter)

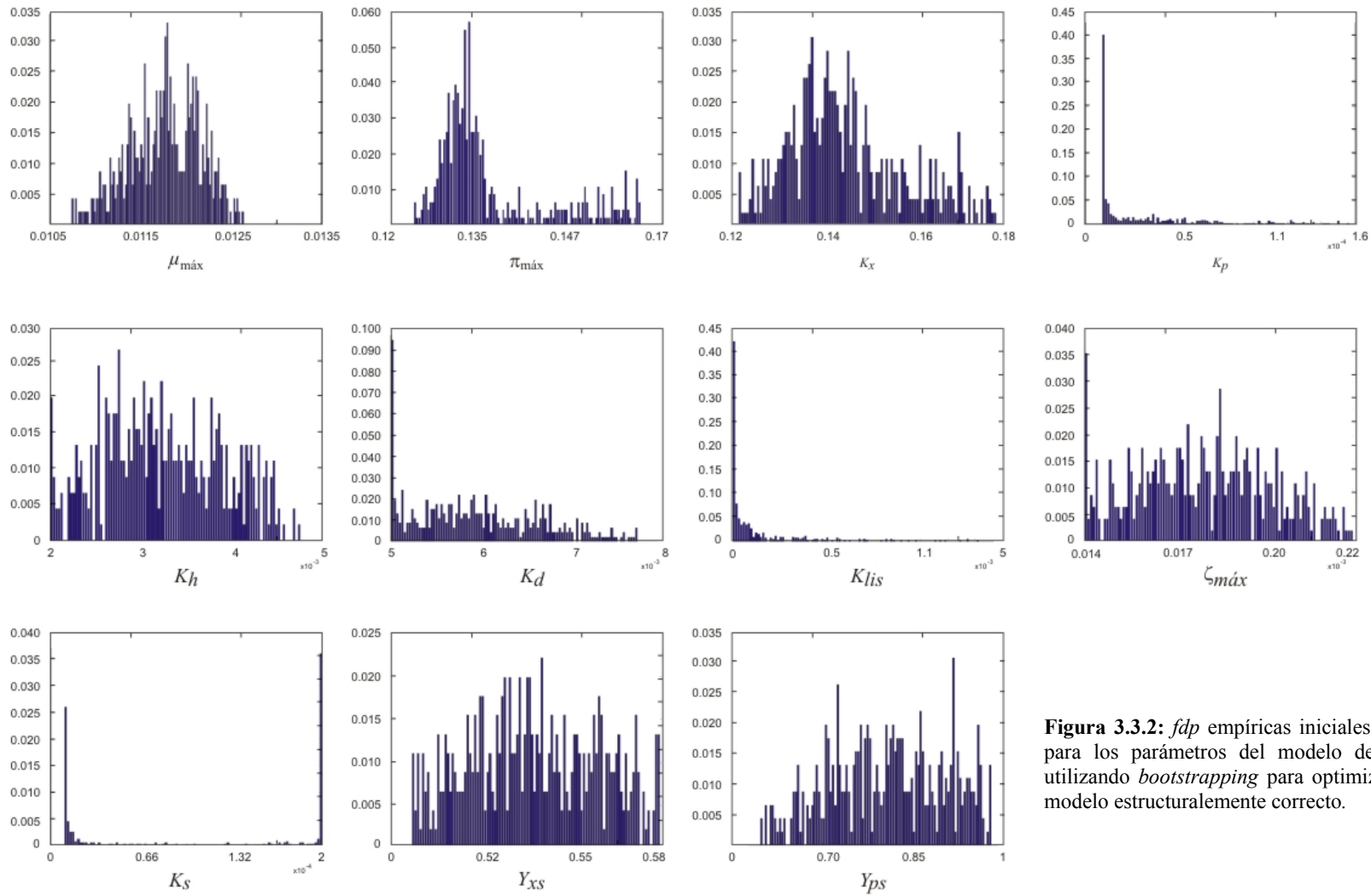


Figura 3.3.2: f_{dp} empíricas iniciales estimadas para los parámetros del modelo de Menezes utilizando *bootstrapping* para optimización con modelo estructuralmente correcto.

Las Figuras 3.3.3-3.3.4 muestran la evolución de las *fdp* durante la optimización del proceso para algunos de los parámetros del modelo. Nuevamente aquí existe un patrón común, que es el grado de acumulación de la *fdp* en una zona muy estrecha del intervalo de variación. Más allá del método de estimación utilizado también podría explicarse porque a medida que los parámetros se conocen más exactamente también hay menor probabilidad de que una interacción entre ellos lleve a un buen ajuste de los datos experimentales. La poca dispersión evidenciada puede ser producto de que algunas combinaciones superaron exitosamente la prueba de *bootstrapping* tanto porque los tiempos de muestreo tenidos en cuenta fueron demasiado poco representativos como así también por alguna influencia del ruido experimental incluido en los datos.

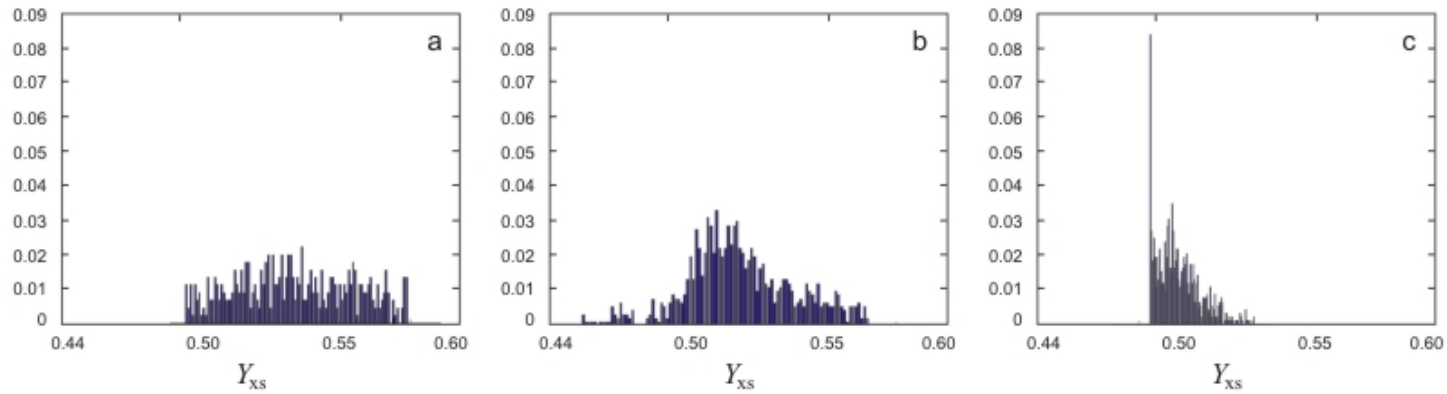


Figura 3.3.3: *fdp* empíricas del parámetro Y_{xs} luego de la mejora del modelo de tendencias con datos de la experiencia 1 (a), 2 (b) y 4 (c)

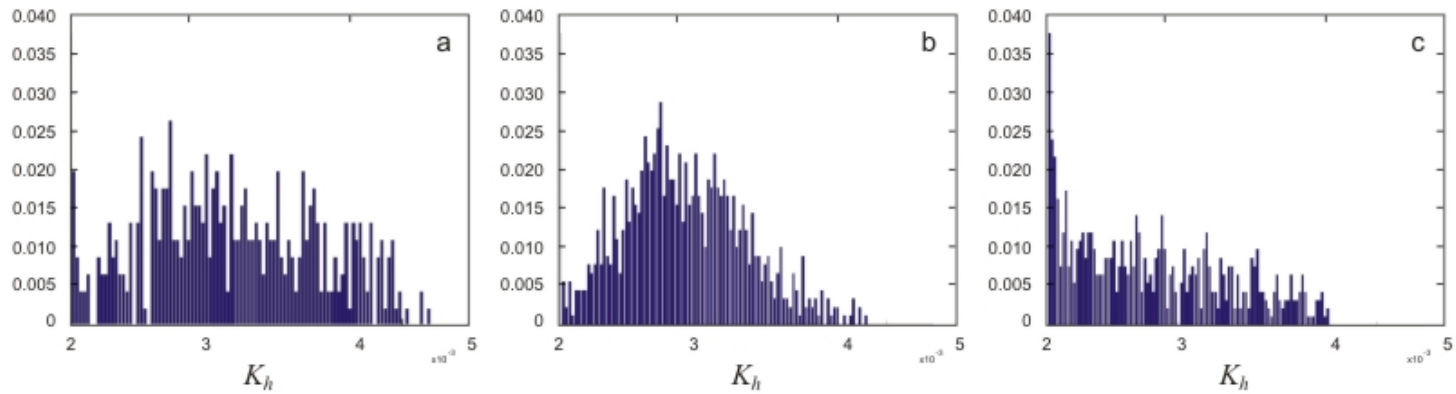


Figura 3.3.4: *fdp* empíricas del parámetro K_h luego de la mejora del modelo de tendencias con datos de la experiencia inicial (a), 1 (b) y 3 (c).

3.4. Diseño de Experimentos Dinámicos para la Optimización de la producción de penicilina en un cultivo de *Penicillium crysogenum* utilizando un modelo de tendencia probabilístico. Proceso estructuralmente diferente al modelo.

En este ejemplo el modelo presentado por Riascos y col. (Riascos y Pinto, 2004) ha sido utilizado como biorreactor *in silico*. El modelo ha sido adaptado paramétricamente para que los intervalos de incertidumbre paramétrica publicados por Menezes y col. (1994) puedan ser útiles.

El biorreactor *in silico* está regido por las siguientes ecuaciones:

$$\begin{aligned} \frac{dV}{dt} &= F_{in} - F_{evap} \\ \frac{dX}{dt} &= \mu X - K_d X - \frac{(F_{in} - F_{evap})}{V} X; \quad \mu = \frac{\mu_{máx} S}{K_x X + S} \\ \frac{dS}{dt} &= -\sigma X + \langle S_{in} F_{in} \rangle - \frac{(F_{in} - F_{evap})}{V} S; \quad \sigma = \frac{\mu}{Y_{xs}} + \frac{\pi}{Y_{ps}} + \zeta; \quad \zeta = \frac{\zeta_{máx} S}{K_s + S} \\ \frac{dP}{dt} &= \pi X - \frac{(F_{in} - F_{evap})}{V} P - K_h P; \quad \pi = \frac{\pi_{máx} S}{K_p + S \left(1 + \frac{S}{K_{in}} \right)} \end{aligned} \quad (3.4.1)$$

y condiciones iniciales:

$$X(0) = 1; S(0) = 4; P(0) = 0$$

donde F_{in} es el caudal de ingreso de nutrientes, F_{evap} es la pérdida de líquido debido a la evaporación del medio de cultivo y S_{in} es la concentración de sustrato en la alimentación del biorreactor.

La Tabla 3.4.1 presenta los valores puntuales de los parámetros del modelo. Los intervalos de los parámetros del modelo de tendencias al comienzo del estudio han sido los mismos que los publicados por Menezes y col. (1994). El óptimo global ha sido calculado con las mismas restricciones operativas que el ejemplo anterior.

La Tabla 3.4.2 muestra la política de operación óptima junto con las políticas diseñadas por la metodología propuesta e implementadas en cada iteración.

Tabla 3.4. 1: Parametrización del biorreactor *in silico*

Parámetro	Símbolo	Unidades	Biorreactor "in silico"
Velocidad máxima de crecimiento específica	$\mu_{m\acute{a}x}$	h^{-1}	0,15
Constante de saturación del crecimiento celular	K_x	g-sustrato/g-DW	0,06
Rendimiento de biomasa respecto de sustrato	Y_{xs}	g-DW/g-sustrato	0,45
Rendimiento de penicilina respecto de sustrato	Y_{ps}	g-Penicilina/g-sustrato	0,9
Velocidad máxima de producción de penicilina	$\pi_{m\acute{a}x}$	h^{-1}	0,012
Constante de saturación de producción de penicilina	K_p	$g L^{-1}$	0,0001
Velocidad máxima de consumo de sustrato para el mantenimiento	$\zeta_{m\acute{a}x}$	h^{-1}	0,025
Constante de saturación para el mantenimiento celular	K_s	$g L^{-1}$	0,0001
Constante de inhibición de producción de penicilina	K_{in}	$g L^{-1}$	0,1
Velocidad de hidrólisis de penicilina	K_h	h^{-1}	0,003

La Figura 3.4.1 ilustra los perfiles de alimentación propuestos en cada experimento para optimización junto con el óptimo calculado. Puede ser observado el cambio en la forma de la convergencia del método respecto al ejemplo anterior cuando no existía incertidumbre estructural. No se producen fluctuaciones respecto al perfil de alimentación óptimo sino que, luego del experimento inicial, las políticas de operación propuestas se acercan en forma gradual y creciente hacia el óptimo. Sin embargo, si bien la metodología produce una mejora significativa en el índice de *performance* del bioprocesos (más del 300%) no es capaz de llegar al óptimo respecto a la función objetivo del proceso y menos aún a la forma de operarlo ya que la convergencia es lograda con una producción 9% por debajo de la máxima posible.

Tabla 3.4. 2: Parámetros de la política de operación durante optimización experiemntal y los de la política de operación óptima

Parámetro	Valor inicial	Límite inferior	Límite superior	Optimo 1ra iteración	Optimo 2da iteración	Optimo 3ra iteración	Optimo 4ta iteración	Optimo "real"
A [L h ⁻¹]	0,6882	0	4,13	0,8707	0,9697	1,3494	1,3036	1,2755
B [h]	0,1431	0,1	0,86	0,1	0,1022	0,1015	0,1000	0,1018
C [h ⁻²]	0,0002	-0,0008	0,0012	2e-4	3e-4	9e-4	8e-4	0,0012
t_0 [h]	0	0	24	24	23,6	24	24	22,37
t_f [h]	240	200	300	300	300	294,8	300	300
<i>Conc. Sustr. Alimentac.</i> [g L ⁻¹]	240	200	500	500	500	500	500	500
<i>Tiempo primera descarga</i> [h]	24	24	48	24	24,17	24,02	24	24
<i>Volumen descarga</i> [L]	60	30	80	80	80	80	80	79,57
<i>Frecuencia descarga</i> [h]	24	24	60	24	24	24,62	24	35,28
<i>Volumen inicial</i> [L]	600	500	700	500	500	500	500	500
<i>Penicilina obtenida, J</i> (Kg)	16,12			35,47	40,49	57,51	53,6	63,24
<i>std(J)</i> (Kg)				3,1	2,1	2,5	1,9	

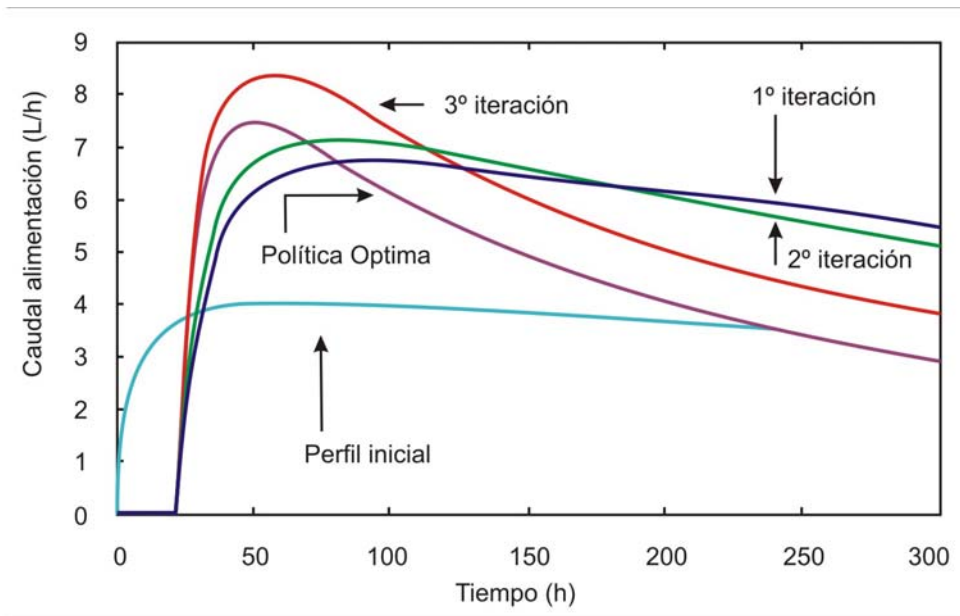


Figura 3.4.1: Políticas de alimentación del biorreactor implementadas durante optimización y la política óptima.

Las diferencias estructurales del modelo respecto al biorreactor *in silico* pueden ser evaluadas al menos cualitativamente observando las discrepancias que existen entre las trayectorias reales y predichas de las variables de estado; principalmente la que corresponde a la concentración de penicilina.

La Figura 3.4.2 ilustra las trayectorias estimadas de las variables de estado junto con los datos experimentales para las iteraciones 0, 1 y 3. Al implementar la política de operación inicial las curvas son diferentes pero hacia el final de la corrida el modelo es capaz de predecir con un error bajo la concentración de producto. Sin embargo, a medida que la optimización experimental es realizada las diferencias se profundizan hasta que finalmente el modelo ya no es capaz de capturar la dinámica de la producción de penicilina. Este resultado llevaría a la evaluación de la posibilidad de mejorar la herramienta de optimización (el modelo de tendencia) en caso de tener que mejorar aún más la performance del proceso estudiado. Esto pone en evidencia el agotamiento en la herramienta de optimización anteriormente observado en el ejemplo 2 del capítulo 2.

Los perfiles de sensibilidad de las estimaciones del índice de performance respecto a los parámetros del modelo fueron calculados a partir de *GSA*. La Tabla 3.4.3 muestra los

correspondientes Si . Los parámetros más sensibles, y por lo tanto re-estimados, están resaltados. En todas las iteraciones para la optimización, la incertidumbre en J estuvo muy concentrada en pocos parámetros. Inversamente a lo que sucedió en el ejemplo anterior, en presencia de incertidumbre estructural en lugar de producirse una dispersión de las fuentes de incertidumbre paramétrica esta fue concentrándose en parámetros individuales (iteraciones 3 y 4).

Tabla 3.4. 3: Índices de sensibilidad para los parámetros del modelo

Parámetro	Si (1ra iter.)	Si (2da iter.)	Si (3ra iter.)	Si (4ta iter.)
$\mu_{m\acute{a}x}$	0	0	0.0329	0
K_x	0.0236	0.0012	0.5910	0.0029
K_d	0.0005	0.0013	0.0015	0.0056
k_{lis}	0.0003	0.0014	0.0384	0
Y_{xs}	0.0341	0.1289	0	0
Y_{ps}	0.3838	0.0012	0.0062	0.0112
$\pi_{m\acute{a}x}$	0.3888	0.0426	0.0191	0.2501
K_p	0.0004	0.0013	0.0015	0
$\zeta_{m\acute{a}x}$	0.0875	0.5224	0.0364	0.0001
K_s	0.0004	0.0014	0.0015	0
K_h	0.0114	0.2627	0.0072	0

Luego de implementada la política de operación inicial y obtenidos los datos experimentales, la estimación de las fdp empíricas de todos los parámetros del modelo han sido estimadas utilizando *bootstrapping* (figuras). La Figura 3.4.3 muestra las fdp y fdp acumulada empíricas de todos los parámetros cinéticos.

Observando las fdp puede decirse que el fenómeno de localización de una alta probabilidad en un rango muy estrecho del intervalo de variación del parámetro es la regla y no la excepción como en el caso anterior, lo que podría deberse más a un problema estructural del modelo que ha cuestiones vinculadas con el método de estimación.

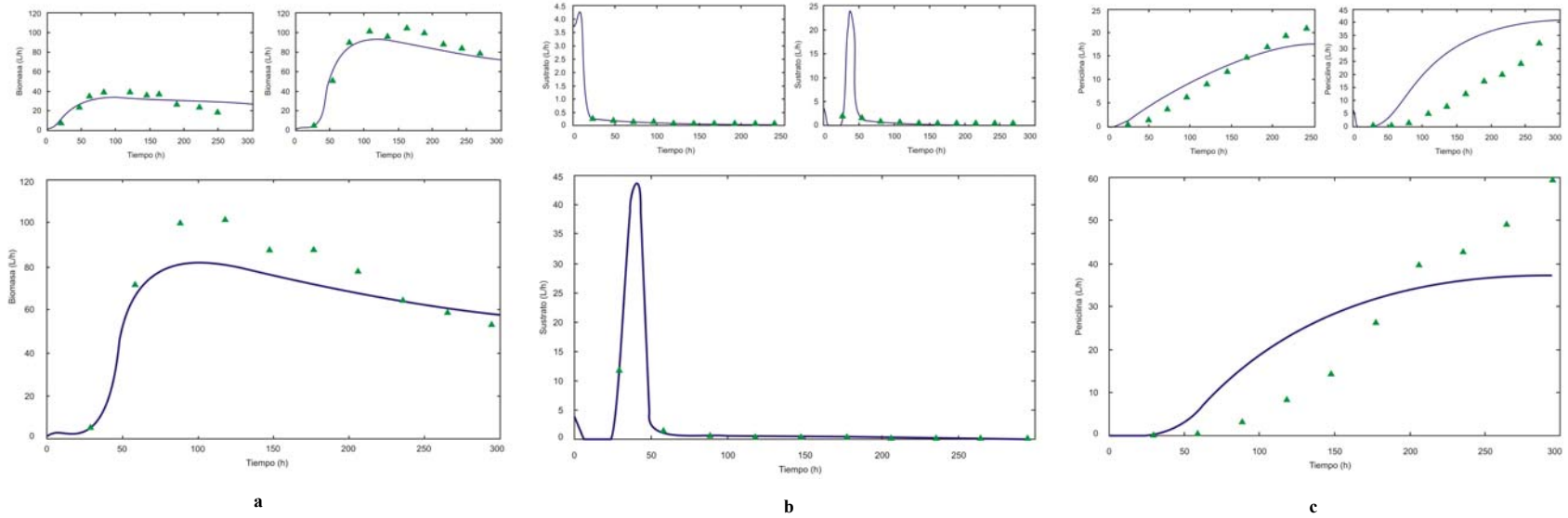


Figura 3.4.2: Trayectorias de las variables de estado (líneas) y datos experimentales (Δ) durante optimización: (a) biomasa; (b) sustrato y (c) penicilina.

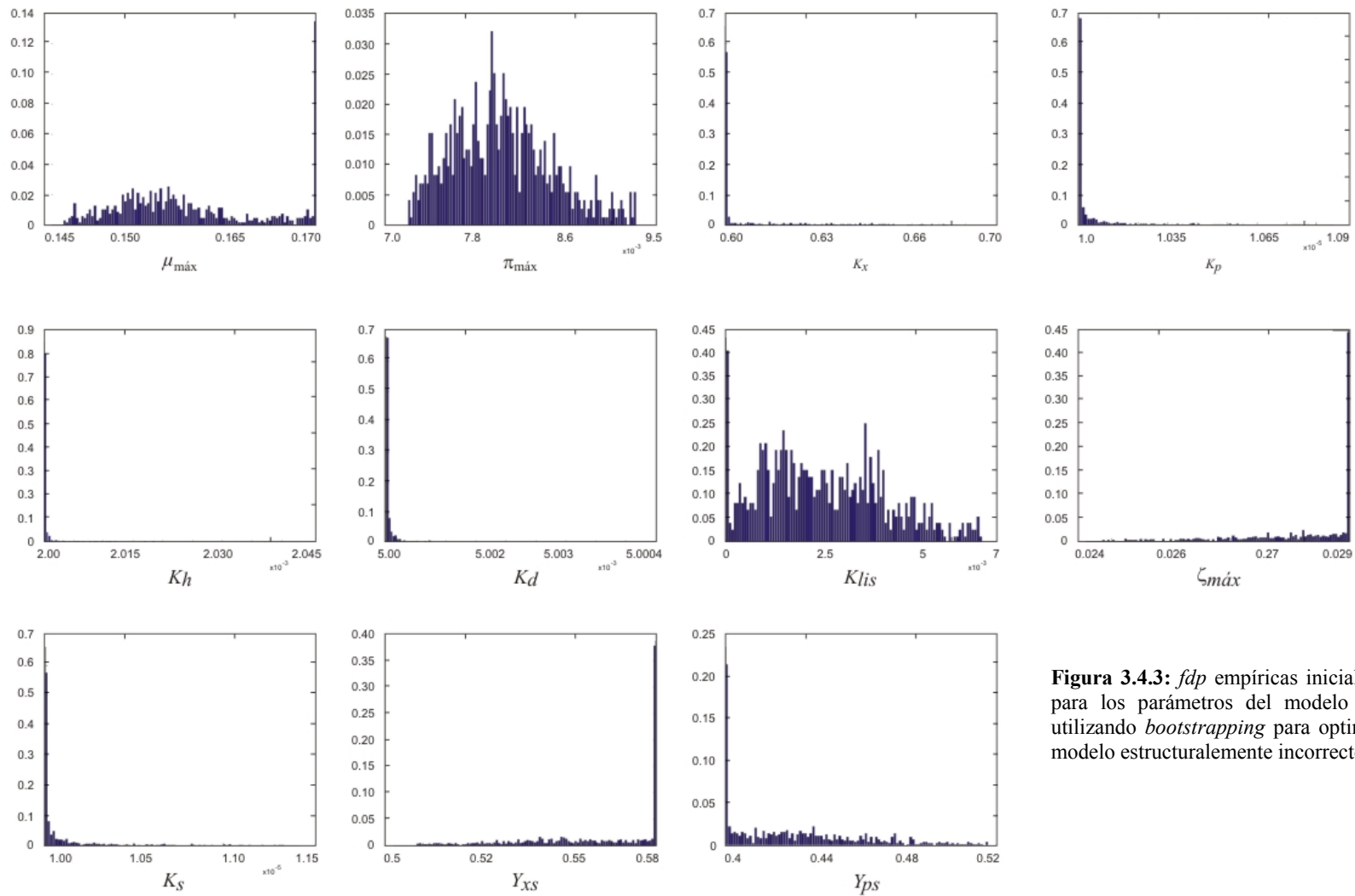


Figura 3.4.3: *fdp* empíricas iniciales estimadas para los parámetros del modelo de Menezes utilizando *bootstrapping* para optimización con modelo estructuralmente incorrecto.

La Tabla 3.4.4 consigna los intervalos para el 95 percentilo de los parámetros más relevantes en cada iteración del método de optimización.

Tabla 3.4. 4: Rangos iniciales y re-estimados de los parámetros del modelo durante la optimización.

Parámetro	Rango inicial	Rango re-estimado
$\mu_{m\acute{a}x}$	0,12 – 0,17	0,1444 – 0,17 (0 iter)
K_x	0,006 – 0,4	(6,0 – 6,89)e-2 (0 iter) (6,0 – 6,89)e-2 (3 iter)
K_d	(5 – 10)e-3	(5 – 5,0004)e-3 (0 iter)
k_{lis}	(0,01 – 8)e-3	1e-6 – 6,65e-3 (0 iter)
Y_{xs}	0,40 – 0,58	0,40 – 0,58 (0 iter) 0,40 – 0,4857 (1 iter) 0,40 – 0,4757 (2 iter)
Y_{ps}	0,4 – 1	0,4 – 0,5178 (0 iter)
$\pi_{m\acute{a}x}$	(3 – 15)e-3	(7,2 – 9,3)e-3 (0 iter) (7,2 – 8,3)e-3 (1 iter) (7,2 – 7,3)e-3 (4 iter)
K_p	(1 – 20)e-5	(1 – 1,09)e-4 (0 iter)
$\zeta_{m\acute{a}x}$	0,014 – 0,029	(2,48 – 2,9)e-2 (0 iter) (2,48 – 2,9)e-2 (2 iter)
K_s	(1 – 20)e-5	(1 – 1,14)e-5 (0 iter)
K_h	(2 – 10)e-3	(2,0 – 2,04)e-3 (0 iter) (2,0 – 2,04)e-3 (2 iter)

La mayor reducción en la incertidumbre paramétrica se da inicialmente cuando las *fdp* son estimadas para todos los parámetros del modelo. En los posteriores ciclos de optimización, cuando son realizadas re-estimaciones de los parámetros más relevantes, lo que principalmente se observa es una modificación tanto del valor de la mediana de la distribución que será el utilizado para el diseño del próximo experimento dinámico; como de la forma de la *fdp* que es utilizada en el cálculo de los *Si*.

Las Figuras 3.4.4-6 muestran esos cambios en la morfología de la incertidumbre paramétrica del modelo para algunos parámetros, en particular para Y_{xs} , $\pi_{m\acute{a}x}$ y $\zeta_{m\acute{a}x}$.

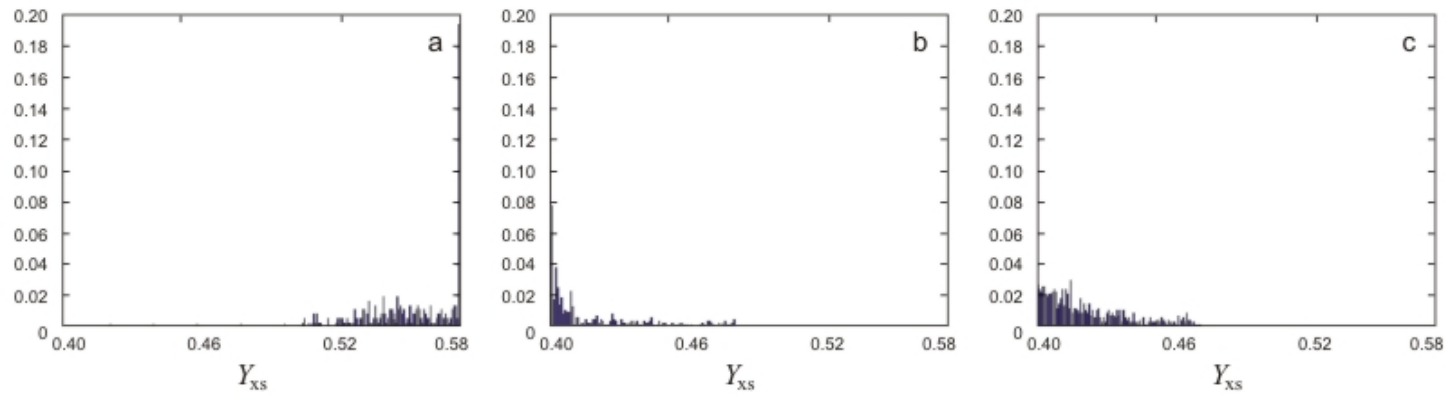


Figura 3.4.4: *fdp* empírica de Y_{xs} inicial (a); iteración 1 (b) y 4 (c).

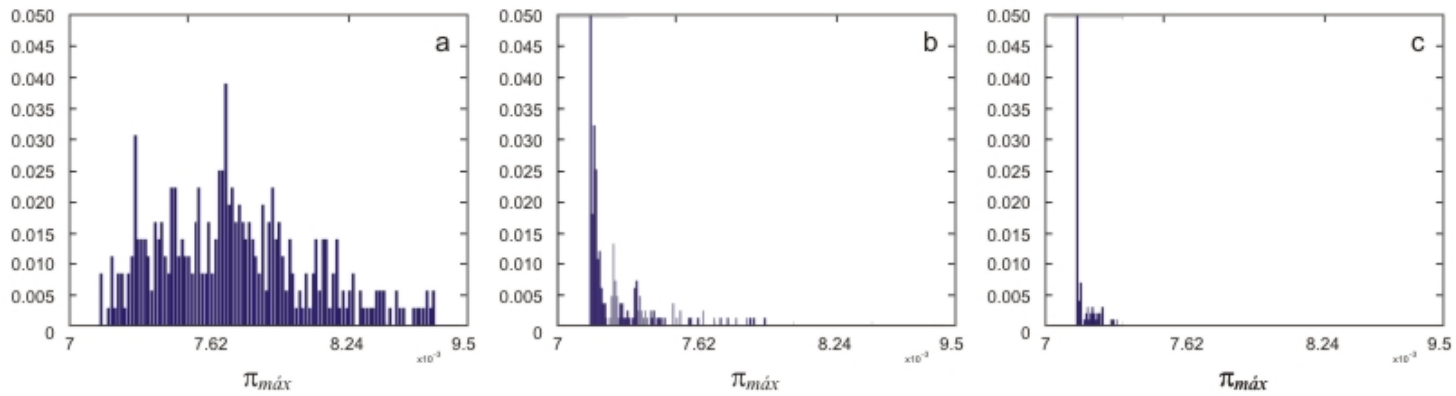


Figura 3.4.5: *fdp* empírica de $\pi_{m\acute{a}x}$ inicial (a); iteración 1 (b) y 4 (c).

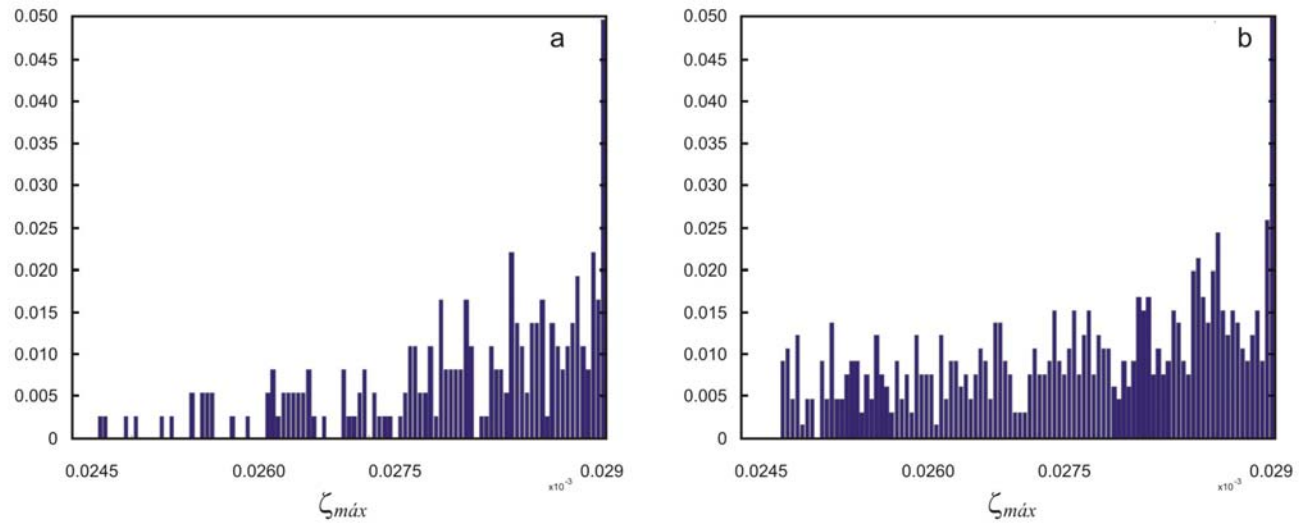


Figura 3.4.6: *fdp* empírica de $\zeta_{m\acute{a}x}$ (a) inicial y (b) iteración 2.

3.5 Conclusiones

La metodología de Diseño de Experimentos Dinámicos basado en Modelos para Optimización ha sido mejorada mediante la incorporación de la técnica de *bootstrapping* en el intento de caracterizar la morfología de la incertidumbre paramétrica del modelo para optimización. Esta ha sido utilizada para la optimización de un biorreactor para la producción de penicilina en cultivo de *P. crysogenum*. El desempeño de la metodología ha sido explorado tanto para el caso en el que se utiliza un modelo con incertidumbre puramente paramétrica (estructuralmente correcto) como cuando la optimización se intenta usando un modelo de tendencias con incertidumbre paramétrica y estructural.

Como conclusiones generales puede decirse que:

- La incorporación de la técnica de bootstrap ha permitido aumentar el conocimiento sobre el espacio paramétrico Θ de los modelos de tendencias ya que permite obtener estimaciones de las *fdp* de los parámetros individuales.
- Durante la optimización experimental ha podido comprobarse la existencia de un sumidero paramétrico y la modificación de las *fdp* hacia intervalos muy estrechos de alta probabilidad. Esto fue evidenciado con mayor nitidez cuando no existe error estructural en el modelo para optimización.
- Aunque puede ser una característica casuística, el modo de convergencia a la política de operación óptima ha mostrado ser diferente para los ejemplos estudiados. Cuando no existe incertidumbre estructural las políticas de alimentación propuestas por el método de optimización oscilan respecto a la política óptima. Cuando el modelo es diferente al proceso fue observado un acercamiento monótono de las políticas propuestas hacia la óptima.

3.6 Bibliografía

- Bajpai, R.K.; Reuss, M. (1980). A mechanistic model for penicillin production. *Journal of Chemical Technology and Biotechnology*; **30**(1), 332-344.
- Chernick, M.R. (1999). *Bootstrap methods, A practitioner's guide*, Wiley, New York.
- Del Barrio, E. Cuestas-Albertos, J.A.; Maltrán, C. (2002). Asymptotic stability of the bootstrap sample mean. *Stochastic Processes and their Applications*, **97**(2); 289-306; 2002.
- De Martini, D.; Rapallo, F. (2008). On multivariate smoothed bootstrap consistency; *Journal of Statistical Planning and Inference*; **138**, 1828-1835.
- Diccio, T.; Efron, B. (1996). Bootstrap Confidence Intervals. *Statistical Science*; **11**; 189-228.
- Dixon, P.M. (2001). The bootstrap and the jackknife: describing the precision of ecological studies, in *Design and Analysis of Ecological Experiments*, 2nd Edition.
- Efron, B.; Tibshirani, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science*; **1**; 54-75.
- El-Nouty, C.; Guillou, A. (2000). On the smoothed bootstrap. *Journal of Statistical Planning and Inference*, **83**; 203-220.
- Gonçalves, S.; White, H. (2004). Maximum likelihood and the bootstrap for nonlinear dynamic models. *Journal of Econometrics*; **119**; 199-219.
- Henderson A. (2005). The bootstrap: A technique for data-driven statistics. Using computer-intensive analyses to explore experimental data. *Clinica Chimica Acta*; **359**; 1-26.
- Kochanski G. (2005). Brute Force as a Statistical Tool; <http://kochanski.org/gpk>.
- Lahiri, S.N.; Furukaga, K.; Lee, Y.-D. (2007). A nonparametric plug-in rule for selecting optimal block lengths for block bootstrap methods; *Statistical Methodology*, **4**(3), 292–321.
- Martin M. (2007). Bootstrap hypothesis testing for some common statistical problems: A critical evaluation of size and power properties. *Computational Statistics & Data Analysis*, **51**; 6321-6342.
- Menezes, J.C.; Alves, S.S.; Lemos, J.M.; Fedo de Azevedo, S. (1994). Mathematical modelling of industrial pilot-plant penicillin-G fed-batch fermentations. *Journal of Chemical Technology and Biotechnology*, **61**(2), 123–138.
- Park, D.; Willemain, T. (1999). The threshold bootstrap and threshold jackknife. *Computational Statistics and Data Analysis*, **31**, 187-202.
- Riascos, C.; Pinto, J. (2004). Optimal control of bioreactors: a simultaneous approach for complex systems. *Chemical Engineering Journal*, **99**(1), 23-34.

¿Qué es la realidad?...

- Mi hijo es un científico.*
- El Tata es zurdito y a veces un poco peroncho.*
- Mi tío es acróbata y malabarista!!!*
- El negro es un jodón bárbaro!*
- Mariano Cristaldi? Ahh...es medio b...*

Sea $f(\bullet)$ la función que describe el comportamiento del agente...

Capítulo 4

***Clonado y Expresión de rBm95 de
Rhipicepalus microplus en Escherichia coli***

4.1. *Rhipicepalus microplus*

4.1.1. Problemática de la garrapata común del ganado

Las garrapatas son ácaros gigantes hematófagos que parasitan externamente a animales superiores (Dale, 2009). *R. microplus* (Figura 4.1) es considerada la garrapata parasitaria más importante en el mundo ya que ataca a ganados tan diversos como el vacuno, de búfalo, de burro, ovino, etc. (FAO, 2003).



Figura 4.1: *R. microplus* hembra (vista ventral)

Puede ser encontrada en regiones tropicales y subtropicales alrededor del mundo. Una representación de los lugares donde potencialmente podría desarrollarse este insecto es mostrado en la Figura 4.2 y debido al fenómeno de calentamiento global esas fronteras se están ampliando con el correr del tiempo.



Figura 4.2: Posibles zonas de desarrollo de *R. microplus*

R. microplus es considerada una endemia en India, las zonas tropicales y subtropicales de Asia, noreste de Australia, Madagascar, sudeste de África, el Caribe, muchos países de América Latina (Informe CFSPH, 2007; Nari, 1996).

Grandes infestaciones (Figura 4.3) con este insecto afectan la producción del ganado ya que cada garrapata hembra que completa su ciclo parasítico ocasiona una disminución del incremento de peso corporal de 0,3 hasta 1,0 g. Esto también repercute en la producción de leche, prolonga el tiempo requerido para que las vaquillonas alcancen el peso apropiado para el primer servicio y disminuye las tasas de preñez. Además puede ocasionar cuadros de anemia más o menos severos o incluso la muerte del animal parasitado. Por otro lado, las heridas derivadas de las picaduras en la piel son potenciales centros de miasis (bicheras) y los cueros obtenidos a partir de animales infestados son de inferior calidad debido a las cicatrices derivadas de las picaduras. Además genera daños ocultos ya que *R. microplus* es el vector de babesiosis (causada por los protozoos *Babesia bigemina* y *B. bovis*) y anaplasmosis (causada por *Anaplasma marginale*) (Willadsen, 1995; 1996; 2001; 2004; Willadsen y Kemp, 1999).



Figura 4.3: Cuello de bovino con infestación aguda de *R. microplus*

Los ambientes propicios para el desarrollo de *R. microplus* son aquellos que presentan temperaturas invernales medias superiores a los 14,5°C y no tienen déficit hídrico. En la Argentina, el área de infestación hasta 2007 tenía como frontera sur más o menos identificada los paralelos 30°-31° en la que más de 10 millones de cabezas eran afectadas observándose una concentración en Corrientes, este del Chaco, este de Formosa y norte de Santa Fe que contienen el 60 % de la población total de bovinos dentro del área infestada con garrapatas. Sin embargo, estudios más recientes estiman que, debido al calentamiento global, los cambios climáticos debidos a la deforestación y factores de movilidad de los rodeos, esa frontera ha sido desplazada hacia el sur entre 3° y 4° (Guglielmone y Mangold, 2007).

4.1.2. Control

La infestación por la garrapata común del vacuno es más abundante en los bovinos de razas europeas (*Bos taurus*) que en los bovinos de razas índicas (*Bos indicus*). Las cruces entre ellos resultan en biotipos con una menor propensión a infestaciones que los bovinos europeos lo que los muestra como alternativa ganadera en regiones infestadas por garrapatas. Sin embargo, en las áreas más favorables para la garrapata, no es posible obtener una ganadería competitiva sin el control de este parásito que es realizado mediante el uso de acaricidas químicos.

El problema que presentan estos agentes de control es la presión selectiva que generan en las poblaciones de insectos que termina en la aparición de cepas resistentes. En Argentina esto ha ocurrido con los acaricidas arsenicales y organofosforados y está habiendo evidencias de poblaciones de garrapatas resistentes a formulaciones basadas en piretroides, lo que conduce a un mayor uso de formamidinas (amitraz) (Cardozo, 1995; Martins, 1995; Romero, 1997).

Es debido a esos problemas de resistencias y que han sido encontrados residuos de pesticidas en carnes, leche y en el medio ambiente que han surgido vacunas contra la garrapata común del ganado. Las principales ventajas de utilizarlas como estrategia de control es que son estables y seguras de transportar, potencialmente económicas y no contaminantes. Sin embargo, su efectividad no es del 100% y aunque han sido reportados importantes disminuciones en el uso de acaricidas, los tratamientos implementados son necesariamente combinados con baños de desinfección.

4.1.3. Antígenos de *R. microplus*. Vacunas

En la actualidad sólo dos empresas en el mundo comercializan formulaciones vacunales contra la garrapata común del ganado: TickGUARD de Hoescht (Australia) y Gavac de Heber Biotech (Cuba). En ellas son utilizados los antígenos Bm86 y Bm95 de *R. microplus*. Las vacunas de primera generación (basada en Bm86) mostraron ser efectivas en diferentes zonas en las que fueron probadas (De la Fuente, 1995; Fragoso, 1990; Rodríguez, 1995; Willadsen, 1996). Sin embargo, en nuestro país no alcanzaron la *performance* requerida para su aprobación identificándose una especie de *R. microplus* resistente a esas vacunas desde la cual ha sido aislado el gen *bm95* a partir del cual ha sido expresada y testada como antígeno efectivo la proteína Bm95.

4.2.1. Descripción fisico-química de Bm95

No existen datos acerca de su ubicación cromosómica ni de la estructura genómica del locus bm95 en el genoma de *Boophilus microplus*.

Bm95 es una glicoproteína que es expresada en la superficie celular de las células epiteliales lumbales del intestino de la garrapata común del ganado bovino *Boophilus microplus* cepa A (cepa argentina resistente a los sueros de animales inmunizados con el antígeno Bm86).

La secuencia nucleotídica de la región codificante (cDNA) ha sido reportada por García-García y col. (García-García, 2000) y ha sido depositada en la base de datos GenBank con número de acceso AF150891. La traducción de dicha secuencia produce un péptido de 569 aminoácidos cuya masa molecular es 63,703 kDa, tanto la composición como la abundancia aminoacídica relativa son mostradas en la Tabla 4.1.

La cantidad de residuos aminoacídicos cargados negativamente (Asp + Glu) serían 80 mientras que los que tendrían una carga positiva (Arg + Lys) sumarían 78. Esto hace que el punto isoeléctrico teórico de Bm 95 (no glicosilada) sea de 6,44.

De su secuencia los softwares libres predicen que su extremo N-terminal es intracelular y que existe una secuencia entre el residuo 6 y 23 que se corresponde con una hélice transmembrana. La proteína contiene en N-terminal una secuencia de aminoácidos hidrofóbicos (19) que es característica de secuencias líderes (Briggs y Gierasch, 1986) y en la posición -1 y -3 hay residuos pequeños y no polares, que también son característicos de péptidos señal. En C-terminal, no contiene una secuencia hidrofóbica que pueda asemejarse a un dominio transmembrana, lo que coincide con las especulaciones para Bm86, en la que se cree que el anclaje se produce mediante glicosil fosfatidilinositol, por lo que la proteína no tendría un dominio citoplasmático.

La evaluación de la distribución de cargas en la estructura primaria de Bm95 no revela la presencia de *cluster* con carga positiva, negativa o altamente cargados indistintamente de la polaridad de la carga.

En base a la secuencia aminoacídica, se estima que la familia de proteínas Bm86, Bm 95, etc., presentan motivos tridimensionales semejantes al EGF los cuales presentan puentes disulfuro entre residuos de cisteínas separados por entre 5 y 7 aminoácidos.

Tabla4.1: Cantidad y abundancia relativa de residuos aminoacídicos en Bm95

Residuo Aminoacídico	Nº total	% abundancia
Ala (A)	32	5.6
Arg (R)	31	5.4
Asn (N)	33	5.8
Asp (D)	34	6.0
Cys (C)	61	10.7
Gln (Q)	20	3.5
Glu (E)	46	8.1
Gly (G)	35	6.2
His (H)	7	1.2
Ile (I)	22	3.9
Leu (L)	29	5.1
Lys (K)	47	8.3
Met (M)	8	1.4
Phe (F)	13	2.3
Pro (P)	25	4.4
Ser (S)	33	5.8
Thr (T)	37	6.5
Trp (W)	6	1.1
Tyr (Y)	20	3.5
Val (V)	30	5.3

La secuencia de Bm 95 identificando los respectivos residuos de cisteína es:

H2N-23-C-7-C-4-C-11-CKC-13-C-5-C-4-C-4-C-9-CVC-9-C-9-C-18-CDC-12-C-4-C-6-C-3-C-10-CC-7-C-9-C-11-C-3-C-8-C-3-C-8-CTC-12-C-8-C-7-C-5-C-8-CEC-10-C-3-C-4-C-6-C-8-CYC-37-C-26-C-24-C-7-C-21-C-16-C-10-C-19-C-5-C-7-C-6-C-8-CQC-10-C-5-C-7-C-6-C-8-CKC-5-C-2-COOH

Por otro lado Bm95 presenta 5 secuencias consenso de N-glicosilación (Asn-X-Ser/Thr) pero al ser expresada en *Pichia pastoris* sólo ha sido glicosilada en 4 de esos sitios. Además, presenta 10 sitios de O-glicosilaciones la mayoría ubicados en la región N-terminal.

4.3. Diseño de clonado de *bm95*

4.3.1. Diseño de la secuencia nucleotídica *bm95* para la expresión de rBm95 en *Escherichia coli*

Se ha obtenido la secuencia aminoacídica de la proteína Bm95 de *Rhipicepalus microplus* (*Boophilus microplus*) del banco de datos GenBank número de acceso AF150891 (Anexo 3, secuencia 1). Tal como ha sido sugerido para la optimización de la producción de proteínas recombinantes (Thiry y Cingolani, 2002), a dicha secuencia se la introdujo en el programa de uso libre ubicado en el sitio web <http://www.entelechon.com> con el cual fue obtenida la secuencia nucleotídica de retrotranscripción con los codones utilizados más frecuentemente por *E. coli* (Anexo 3, secuencia 2 y análisis 1).

Esa secuencia 2 ha sido optimizada con el programa Leto 1.0 para su expresión en *E. coli* pero tomando como criterios de optimización no sólo la frecuencia de uso de codones sino que también:

- Minimizar la repetición de codones en tandem.
- Maximizar los codones más utilizados por *E. coli*.
- Minimizar la formación de estructuras secundarias del mRNA.
- Minimizar la repetición de porciones largas en la secuencias.
- Minimizar la cantidad de sitios de restricción.
- Minimizar largas porciones de AT y GC (aunque este parámetro no es importante en la expresión de proteínas si lo es en los pasos de PCR).

Cabe señalar que dicho software no asegura la unicidad de la secuencia reportada por lo que se han hecho varias optimizaciones y ha sido escogida la mejor secuencia cuyo análisis de codones fue el mejor. La secuencia escogida es mostrada en el Anexo 3 como secuencia 3 y el análisis de codones correspondiente es el análisis 2 del Anexo 3.

En dicho análisis puede ser observado que dentro de los primeros diez (10) codones aparecían cuatro (4) con frecuencias de utilización menores a la óptima en *E. coli* en las posiciones: cinco (5) [GCA], siete (7) [TTC], ocho (8) [GTT] y diez (10) [GCA]; por este motivo se optó por cambiarlos por los más utilizados por este microorganismo. Durante el proceso de optimización ha sido generado un sitio de reconocimiento de la endonucleasa EcoRI (5'-GAATTC-3') el cual ha sido mantenido en la secuencia como

posible herramienta de reconocimiento del inserto durante las actividades posteriores de biología molecular.

De la secuencia optimizada también ha sido eliminado el codón de finalización de la traducción TAA y se le han adicionado dos sitios de reconocimiento de enzimas de restricción en los extremos de la secuencia optimizada:

- En 5' y compartiendo el codón de inicio de la traducción el sitio para NdeI (5'-CATATG-3').
- En 3' el sitio para HindIII (5'-AAGCTT-3').

La secuencia finalmente sintetizada es la secuencia 4 del Anexo 3, los sitios de reconocimiento de las endonucleasas aparecen en cursiva y el análisis de los codones es el análisis 3 del Anexo 3.

4.3.2. Sistema de expresión utilizado. Breve descripción

El sistema escogido para la expresión de rBm95 es el formado por *E. coli* BL21(DE3) y los plásmidos 24b y 26b de la serie pET que permiten el clonado en fase de la secuencia sintetizada y además agregan una cola de histidinas en C-terminal de la proteína traducida.

Como características generales de esos componentes puede decirse que:

En *E. coli* BL21(DE3) las siglas indican sus particularidades genotípicas:

- La B es para distinguirla de la cepa K-12 (utilizada anteriormente en biología molecular) ya que presenta ventajas en cuanto a la robustez del crecimiento, lo que potencialmente la hace un mejor sistema de expresión. Además esta cepa es deficiente en las proteasas lon (citoplasmática) y omp (presente en el espacio periplasmático) lo que aumenta la vida media de las proteínas producidas dentro de la célula y además evita su degradación en los siguientes pasos de purificación.
- La sigla L21 significa que la cepa es resistente a la infección con el bacteriofago 21.
- DE3 significa que son lisógenos crípticos de λ DE3 y por lo tanto poseen integrado en su cromosoma una copia del gen de la T7 RNA polimerasa bajo el control del promotor lacUV5 pero que el lisógeno es incapaz de ingresar en fase vegetativa y lisar a la célula. La T7 RNA polimerasa es una transcriptasa extremadamente procesiva y además reconoce una secuencia promotora mínima

y diferente a las reconocidas por las transcriptasas de *E. coli*, lo que asegura que una vez sintetizada solo transcribirá el gen que se encuentre bajo la regulación de su promotor. Por otra parte, el promotor lacUV5 es una versión modificada del promotor lac salvaje, cuya actividad basal es mucho menos sensible a los niveles de cAMP (cAMP desreprime la expresión de los genes corriente abajo del promotor lac), esta característica es importante ya que los niveles de cAMP intracelular se correlacionan con la densidad bacteriana en los cultivos por lo que este promotor asegura la obtención de cultivos con alta carga bacteriana sin la síntesis de proteína recombinante en estadios previos a la inducción.

Los rendimientos en proteínas obtenidos son de aproximadamente 10 – 15 %(w/w) respecto al peso húmedo de bacterias cosechadas para proteínas solubles y del 27 – 33 %(w/w) cuando la proteína se obtiene en cuerpos de inclusión.

Para el caso los vectores del grupo pET puede decirse que ellos son reconocidos como el sistema más eficiente para la expresión de proteínas recombinantes en cepas de *E. coli* clasificadas como DE3 las cuales permiten que las señales de transcripción y traducción (opcional ya que solo se presenta en los vectores pETs que contienen el péptido líder del bacteriófago T7) se encuentren bajo el control de señales utilizadas por el bacteriófago T7.

Los genes heterólogos se clonan en el vector bajo el promotor reconocido por la T7 RNA polimerasa la cual es tan selectiva, activa y eficiente que en cultivos inducidos completamente logra que casi toda la maquinaria involucrada en la síntesis proteica se comprometa con la expresión del gen de interés obteniéndose así rendimientos de proteína recombinante mayores al 50% respecto al total de proteína bacteriana luego de pocas horas de inducción. Además el nivel de expresión puede controlarse con la concentración de inductor que se utiliza.

Otra ventaja del sistema es que la producción de la proteína recombinante solo se produce bajo la cascada de señales disparadas por el inductor que en el caso de la cepa escogida es IPTG o lactosa.

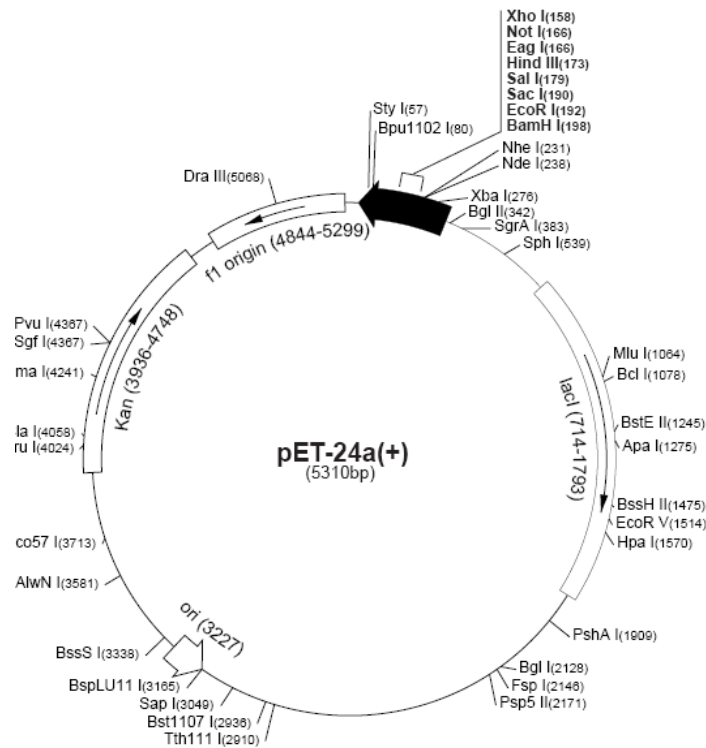


Figura 2.4: Representación del vector de expresión pET-24
<http://www.merckbiosciences.com/docs/NDIS/TB072-000.pdf>

Una representación pictórica del vector de expresión pET-24 es mostrada en la Figura 4.4. Este vector deriva del vector pBR322 ya que presenta su origen de replicación y contiene las siguientes características:

- Gen de selección que otorga resistencia al antibiótico kanamicina a las bacterias transformadas.
- Gen *lacI* que es el regulador natural del operón *lac*, este actúa en *trans* y regula la expresión de la T7 RNA polimerasa [cromosómica en las bacterias (DE3)] y de la proteína de interés debido a su interacción con el promotor *T7lac* al unirse a la secuencia *lacO* corriente abajo de la secuencia promotora T7.
- El promotor *T7lac* es utilizado para controlar la expresión basal de la proteína heteróloga, la cual puede ser tóxica y producir desde disminución en la velocidad de crecimiento hasta la muerte celular. Éste evita la traducción por parte de la T7 RNA polimerasa ya que contiene la secuencia *lacO* que actúa en forma *cis* permitiendo la unión del producto de *lacI* hasta que se induce la síntesis de la T7 RNA polimerasa.

- Sitio de unión a ribosoma (RBS) altamente eficiente derivado del mRNA de la proteína mayor de la cápside del fago T7 lo que asegura un buen anclaje del sistema de traducción al mRNA.
- Secuencia de agrega una cola de poli-Histidina (6 His) en la región C-terminal de la proteína de interés lo que facilita la purificación del producto mediante cromatografía de unión a metales y el reconocimiento del mismo por ensayos de western blot en los que son utilizados anticuerpos específicos.
- Secuencia de finalización de la transcripción del T7 (T7 terminator) para evitar la transcripción inespecífica de mRNA.

El funcionamiento del sistema de expresión es mostrado en la figura 4.5.

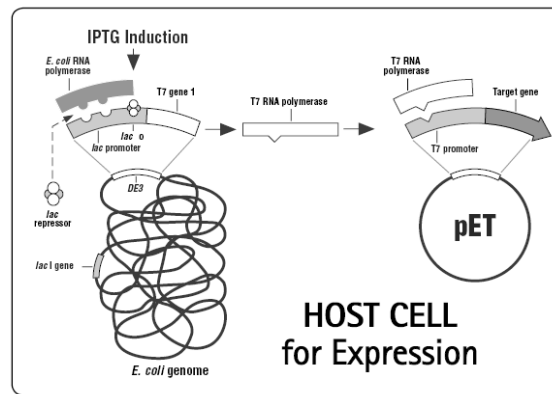


Figura 3.5: Funcionamiento general del sistema DE3-pET.
(pET System Manual, Novagen)

4.3.3. Clonado y expresión de rBm95 en *E. coli*

Ha sido recibida una construcción genómica de la empresa GenArt SA correspondiente a un plásmido cerrado el cual contenía la secuencia nucleotídica optimizada (geneCRISTALDI en la Figura 4.6) que codifica para la proteína Bm95 de *Boophilus microplus* (secuencia 4 del Anexo 3). El mapa del vector de clonado es mostrada en la Figura 4.6. El vector completo tiene 4007 bp y consta de un origen de replicación para su clonado en *Escherichia coli*, la secuencia nucleotídica de un gen que confiere resistencia a kanamicina a las células transformantes, el geneCRISTALDI (1732 bp) que contiene los sitios reconocidos de las endonucleasas NdeI (378), EcoRI (467) y HindIII (2087).

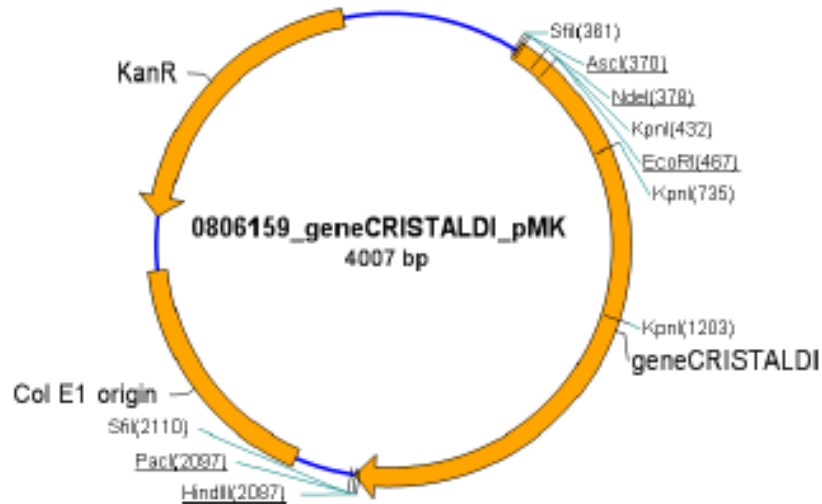


Figura 4.6: Mapa del plásmido proporcionado por la empresa GenArt SA

4.3.4. Clonado de construcción geneCRISTALDI_pMK

Los 10 µg de plásmido provistos por GenArt SA han sido resuspendidos en 50 µL de agua ultrapura estéril y 0,5 µL de esa suspensión han sido utilizados para transformar la cepa de clonado *E. coli* Top10 la cual presenta una elevada frecuencia de transformación (10^9 UFC/µg pUC 18) por lo que es normalmente utilizada como instancia de replicación de construcciones genéticas para así obtener grandes cantidades de plásmido de buena calidad que luego será utilizado en las transformaciones de las cepas de expresión y en los controles de inserto y secuenciación (ver Anexo 4, sección A.4.2).

En la Figura 4.7 puede ser observado un gel de agarosa 1% resultante de correr mini preparaciones de DNA plasmídico obtenidas utilizando el kit Wizard Plus de PROMEGA según sus especificaciones. En la calle 1 aparecen los marcadores de MW (ladder 1Kb), en la calle 4 la construcción adquirida de GenArt (\approx 4 Kbp) y en la calle 5 el vector pET-24b+ (\approx 5.3 Kbp) el cual previamente había sido clonado y purificado con el mismo kit.

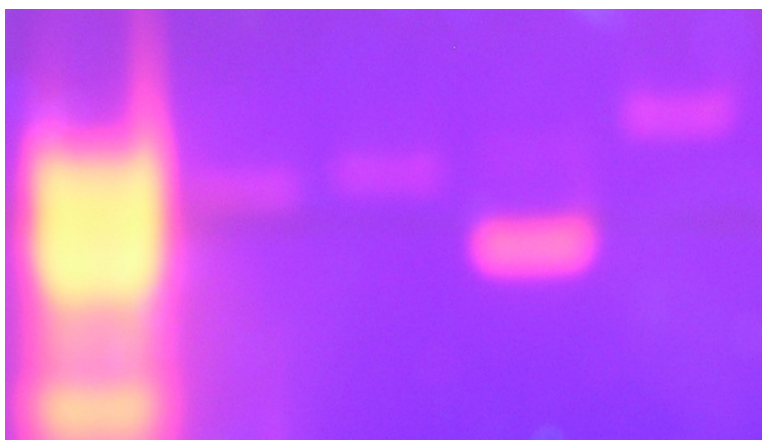


Figura 4.5: Gel de agarosa 1% de minipreps de construcción GenArt (calle 4) y pE T-24b+ (calle 5).

4.3.5. Digestión y Fosforilación

Al corroborar que los plásmidos geneCRISTALDI_pMK y pET-24b+ se correspondían con los MW esperados se procedió a la digestión doble de cada uno con las enzimas de restricción NdeI y HindIII en un buffer compatible para el 100% de eficiencia.

Una vez finalizadas las digestiones (ver Anexo 4, sección A.4.4), los fragmentos cuyos MW coincidían con geneCRISTALDI y pET-24b+ han sido purificados utilizando el kit Wizard Plus de PROMEGA según sus especificaciones.

Debido a que en un primer intento de ligación no se logró obtener una fracción mayoritaria de la construcción pET-24b-geneCRISTALDI, el fragmento correspondiente a geneCRISTALDI fue fosforilado en el extremo 5' con la enzima T4 kinasa con el propósito de aumentar la eficiencia de ligación.

4.3.6. Ligación

La reacción de ligación de los fragmentos digeridos geneCRISTALDI (fosforilado, ver Anexo 4, sección A.4.6) y pET-24b+ ha sido realizada utilizando la enzima T4 ligasa.

4.3.7. Clonado de construcción pET-24+-geneCRISTALDI

La mezcla de la reacción de ligación ha sido utilizada para la transformación de las cepas de *E. coli* Top10. De las colonias transformantes han sido realizadas mini preparaciones de ADNp de diez de ellas seleccionadas al azar a las cuales se las sometió a digestiones con las enzimas de restricción NdeI y HindIII en buffer compatible, de esas digestiones sólo una ha mostrado la liberación de un fragmento cuyo MW era coincidente con el del fragmento geneCRISTALDI (Figura 4.8).

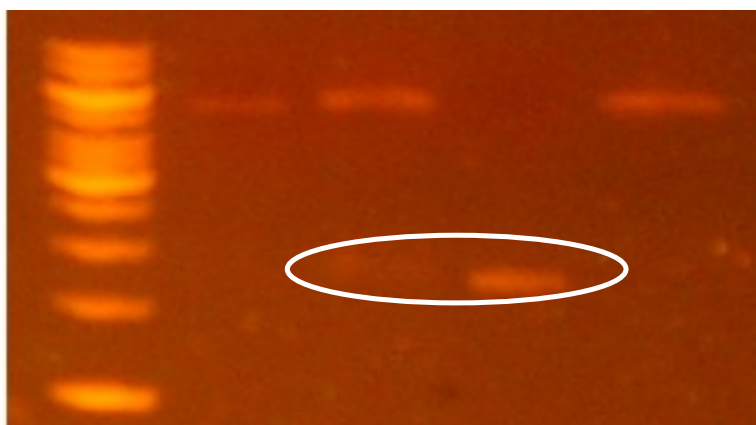


Figura 4.8: Gel de agarosa 1% de minipreps de colonias de *E. coli* Top10 transformantes digeridas: MW ladder (calle 1), colonia negativa (calle 2), colonia positiva (calle 3), fragmento geneCRISTALDI aislado (calle 4), pET-24+ digerido con NdeI y HindIII (calle 5).

4.3.8. Secuenciación de la construcción pET-24+-geneCRISTALDI

La construcción pET-24+-geneCRISTALDI ha sido secuenciada por Ibiotec, INTA Castelar. En una primera iteración la secuencia del fragmento geneCRISTALDI se intentó dilucidar utilizando solamente los cebadores T7 promoter y T7 terminator que flanquean el sitio de múltiple clonado del vector pET-24+. Sin embargo, la baja calidad de resolución de la secuencia en bases alejadas de las zonas de inicio de la replicación impidió la identificación completa del inserto (los datos no se muestran). Entonces se diseñó un cebador que acoplase el la zona media del fragmento a resolver para así obtener la secuencia completa, esto ha sido realizado con la ayuda del software libre Primer-BLAST y la secuencia del cebador óptimo junto con los indicadores son mostrados en la Tabla 4.2.

Tabla 4.2: Secuencia y parámetros fisico-químicos del cebador diseñado

Secuencia	Longitud	%GC	Tmelting
5'- TCTCGTTGTTGTCAGGGTTG-3'	20 bases	50	59,72

Las secuencias obtenidas utilizando los cebadores T7 promoter, T7 terminator y el cebador diseñado por nosotros son mostradas junto con los resultados de los alineamientos realizados en BLAST en el Anexo 3 (secuencias 5, 6 y 7 y alineamientos 1, 2 y 3 respectivamente). Como puede ser observado en todos los casos las secuencias traducidas se correspondían con diferentes fragmentos de Bm95.

La secuencia completa del inserto geneCRISTALDI ha sido reconstruida desde los fragmentos anteriormente estudiados (secuencia 8 del Anexo 3), en ella han sido resaltados los sitios de reconocimiento de las enzimas de restricción NdeI, HindIII y EcoRI (en negrita y cursiva), la secuencia del cebador de la Tabla 2 (en negrita y subrayado), la secuencia que codifica para la cola de polihistidina en fase (en negrita, cursiva y subrayado) y el sitio de finalización de la traducción (en negrita y rojo). El alineamiento de la secuencia reconstruida en BLAST resultó coincidente en un 99% con la secuencia aminoacídica de Bm95 (alineamiento 4 del Anexo 3) corroborándose así la identidad.

4.3.9. Expresión de Bm95

La construcción pET-24b+-bm95 ha sido utilizada para transformar células de *E. coli* BL21(DE3). Colonias transformantes han sido elegidas al azar y fueron realizados ensayos de expresión de Bm95 utilizando caldo LB como medio de cultivo e IPTG⁸ en una concentración final de 1 mM como inductor. Las fracciones totales de proteínas celulares han sido analizadas mediante SDS-PAGE y se ha detectado la sobreexpresión de una proteína cuyo MW que se corresponde con Bm95. Una de las colonias positivas ha sido cultivada en dos medios de cultivo (caldos LB y TB) selectivos para observar si existían diferencias en la sobreexpresión con IPTG 1mM. El criterio de normalización para la comparación ha sido la cosecha de un volumen constante de medio de cultivo (1 mL), además han sido tomadas muestras a diferentes tiempos. Los resultados son mostrados en la Figura 4.9 en la que puede ser observada:

- La presencia de la banda de sobreexpresión coincidente con el MW de Bm95.
- La diferencia de intensidad de las bandas producto de un mayor crecimiento celular en el caldo TB, a partir de lo que se decidió realizar los ensayos posteriores utilizando ese medio de cultivo.
- La intensidad de la señal de sobreexpresión, en ambos medios de cultivo, es menor cuando la inducción es realizada durante toda la noche.

⁸ IPTG: siglas que representan β -D-tiogalactósido de isopropilo

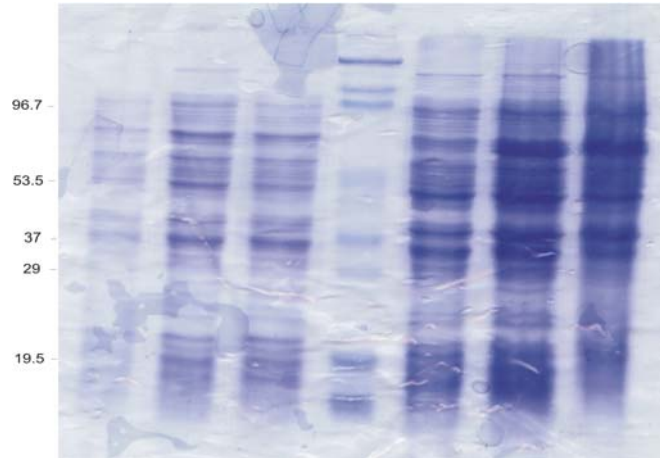


Figura 4.6: SDS-PAGE revelado con Coomassie Brilliant Blue de cultivos de *E. coli* BL21 transformada con la construcción pET-24b-*bm95* : calle 1, 2 y 3: cultivo en caldo LB sin inducir, luego de 3 horas de inducción, inducido toda la noche, respectivamente; calle 4: MW ladder; calle 5, 6 y 7: cultivo en caldo TB sin inducir, luego de 3 horas de inducción, inducido toda la noche, respectivamente.

Para corroborar la identidad de la banda de sobreexpresión ha sido realizado un nuevo ensayo de expresión en caldo TB, SDS-PAGE seguido de western blot utilizando un anticuerpo anti-Histag acompañado a un fluoróforo cuya presencia es detectada mediante irradiación UV. En la Figura 4.10 es mostrada la presencia de fluorescencia de

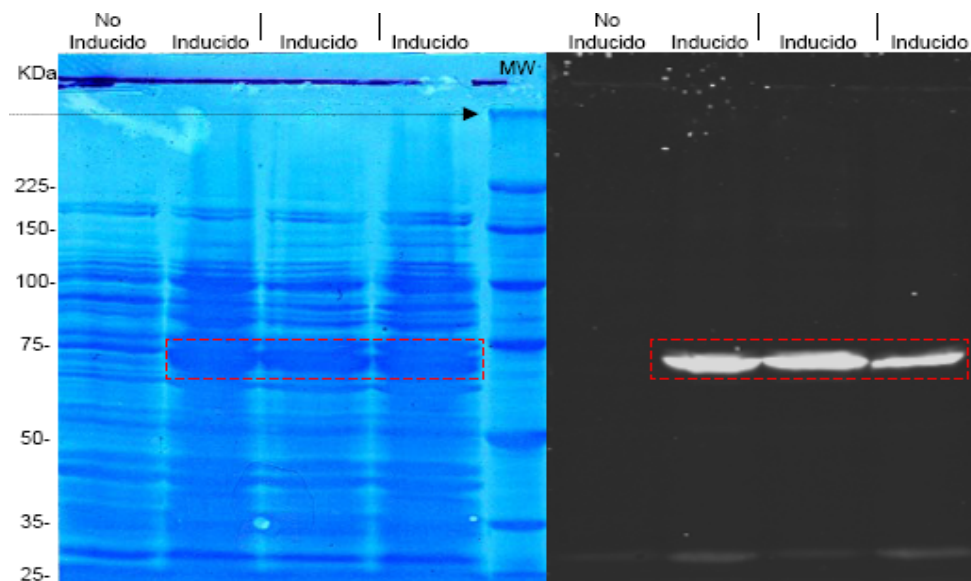


Figura 4.7: SDS-PAGE revelado con Coomassie Brilliant Blue (calles 1-5) y Western Blot con anti-Histag irradiado con luz UV de cultivos de *E. coli* BL21 transformada con la construcción pET-24b-*bm95* : calles 1 y 6: sin inducir; calles 2-4 y 7-10: inducidos.

una banda cuyo MW coincide con el de Bm95-Histag por lo que esto ha sido tomado como prueba definitiva de la identidad proteica del producto de sobreexpresión como Bm95.

Con el objetivo de determinar en qué fracción celular se encuentra rBm95 se ha realizado un experimento de inducción (IPTG 0.5 mM, 30°C). A las muestras recolectadas a intervalos regulares de 1 hora se las sometió a lisis celular y separación de la fracción de proteína soluble e insoluble [centrifugación a 25.000 rpm, 15 minutos: sobrenadante (fracción soluble); pellet (fracción insoluble)] para luego hacer un SDS-PAGE.

En la Figura 4.11 puede observarse que tanto en las fracciones insolubles como solubles del cultivo inducido aparece una banda que se corresponde con el MW de rBm95 siendo más intensa en el primer grupo. Esto llevaría a pensar que la proteína heteróloga se encuentra mayoritariamente como cuerpo de inclusión siendo que a medida que transcurre el tiempo la porción soluble disminuye en relación a la que se encuentra insoluble. Sin embargo, en la fracción insoluble del cultivo no inducido también aparece una banda que podría corresponderse con rBm95 no siendo así en la otra fracción.

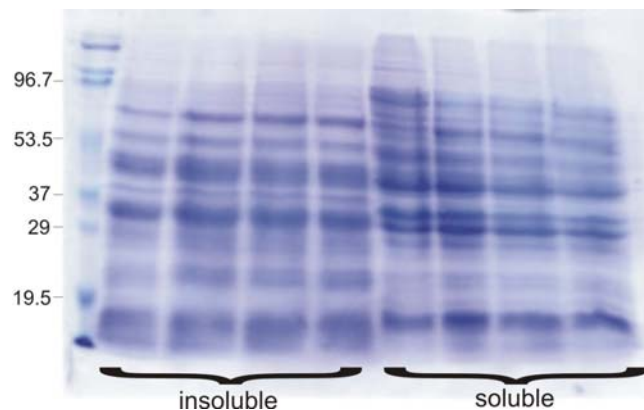


Figura 4.8: SDS-PAGE revelado con Coomassie Brilliant Blue de cultivos de *E. coli* BL21 transformada con la construcción pET-24b-*bm95* : calle 1: MW ladder; calle 2-5: fracción insoluble de cultivo sin inducir y 1, 2 y 3 hs post inducción, respectivamente; calle 6-9: fracción soluble de cultivo sin inducir y 1, 2 y 3 hs post inducción, respectivamente.

Para intentar dilucidar si la banda observada en el cultivo no inducido era producida por una expresión basal de rBm95 se realizó una mini purificación con perlas de agarosa aprovechando la cola de poli-histidina en C-terminal adicionada a rBm95 (ver Anexo 4, sección A.4.8).

En la Figura 4.12 puede observarse que una proteína de aproximadamente 65 kDa sólo estuvo presente en las fracciones solubles e insolubles de los cultivos inducidos lo que se tomó como indicativo de la presencia de rBm95.

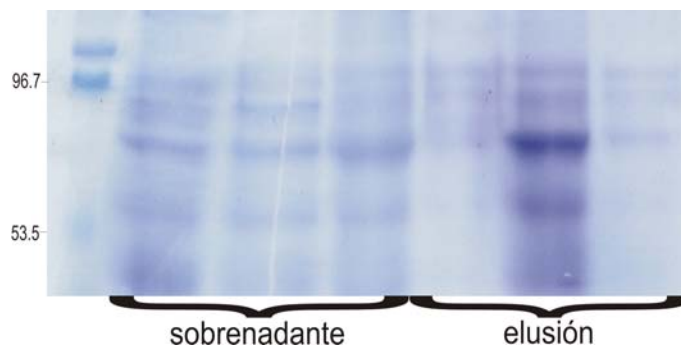


Figura 4.9: SDS-PAGE revelado con Coomassie Brilliant Blue de cultivos de *E. coli* BL21 transformada con la construcción pET-24b-*bm95* : calle 1: MW ladder; fracción insoluble de cultivo sin inducir: sobrenadante (calle 2) y elusión (calle 5); fracción insoluble de cultivo inducido (4h): sobrenadante (calle 3) y elusión (calle 6); fracción soluble de cultivo inducido (4h) : sobrenadante (calle 3) y elusión (calle 6).

4.3.10. Análisis de estabilidad de la construcción pET-24+*-bm95* en *E. coli* BL21(DE3)

La estabilidad de una construcción genética es medida como el porcentaje de la población celular de un cultivo que luego de cierto período de tiempo de haber crecido en un medio no selectivo ha retenido la construcción. Para evaluar la estabilidad de la construcción genética pET-24+*-bm95* han sido realizado ensayos de recuento en placa (Anexo 4, sección 5.9) en medios selectivos (kanamicina 30 mg.L⁻¹) y no selectivos. Los resultados son mostrados en la Tabla 4.3 en la que T1 corresponde al tiempo al momento en el que el cultivo alcanzó una OD₆₀₀ de 1, las muestras han sido tomadas y los cultivos inducidos con IPTG 1 mM. T2 corresponde a los recuentos de las muestras tomadas 8 horas postinducción.

Como puede ser observado el cultivo de las células transformadas con pET-24+*-bm95* en caldo no selectivo presenta un alto porcentaje de estabilidad, la que disminuye luego de un largo período post-inducción de la expresión de la proteína recombinante. Esos resultados son coincidentes con los reportados por Panda (Panda y col., 1999). Un dato que debe ser considerado es que aunque los recuentos disminuyeron en un orden de magnitud, las OD₆₀₀ de los cultivos fueron mayores a T2 lo que podría indicar que la esa disminución en el porcentaje de supervivencia esté afectado tanto por la pérdida efectiva

de la construcción (y por lo tanto la resistencia al antibiótico) como por la pérdida de viabilidad de aquellas células con alto contenido de proteína heteróloga.

Tabla 4.3: Estabilidad de la construcción pET-24+-*bm95* en *E.coli* BL21(DE3)

		UFC mL ⁻¹ x 10 ⁻⁸
T1	Kan (-)	65
	Kan (+)	62
	Estabilidad	95%
T2	Kan (-)	9,5
	Kan (+)	7,2
	Estabilidad	76%

4.4 Conclusiones

Los efectos de infestaciones con la garrapata *Rhipicepalus microplus* en la producción pecuaria son conocidos alrededor del mundo. En el intento de combatir los problemas vinculados con las pérdidas económicas han sido desarrolladas vacunas contra ese insecto basadas en las proteínas Bm86 y Bm95, siendo esta última encontrada en el intestino de la cepa argentina que presenta resistencia a las formulaciones basadas sólo en Bm86.

Los objetivos cumplidos en este capítulo son:

- La secuencia aminoacídica de Bm95 ha sido caracterizada y un fragmento de ADN codificante para *bm95* cuya secuencia ha sido optimizada para su expresión en *E. coli* ha sido sintetizado, clonado en dicho organismo.
- Las pruebas de inducción de la expresión de rBm95 han sido exitosas ya que ha podido ser identificada una banda de sobreexpresión que coincide con la masa molecular teórica de rBm95 que presenta una cola de poli-histidina. Además, luego de 4 horas de inducción de la expresión la fracción mayoritaria de rBm95 se encuentra en la fracción insoluble del lisado de *E. coli*.
- La construcción genética pET-24+-*bm95* demostró ser estable en *E.coli* BL21(DE3).

4.5 Bibliografía

Boué O, Farnós O, González A, Fernández R, Acosta JA, Valdés R, González LJ, Guanche Y, Izquierdo G, Suárez M, Domínguez I, Machado H, Rodríguez M, Leonart R. (2004). Production and biochemical characterization of the recombinant *Boophilus microplus* Bm95 antigen from *Pichia pastoris*; *Experimental and Applied Acarology*, **32**, 119-128.

Briggs M. and Gierasch L. (1986). Molecular mechanisms of protein secretion: the role of the signal sequence; *Advances in Protein Chemistry*. Academic Press Inc. **38**, pp. 109–180.

Cardozo, H. (1995). Situación de resistencia del *Boophilus microplus* en Uruguay. En: “III Seminario Internacional de Parasitología Animal”. (Rodríguez C., S. & Frago S., H., ed.). Acapulco. México. 11-13, Octubre: 30-38.

Dale W.; Plagas y Pestes Médico Veterinaria. Garrapatas neotropicales; 2009. (http://www.lamolina.edu.pe/profesores/wdale/ent_med_vet/2/PLAGAS%20PESTS%20M%C9DICAS%20VETERINARIAS.%20GARRAPATAS%20NEOTROPICALES.%20VERSI%D3N%2001.T09.%20WILLIAM%20E.%20DALE%20PHD..pdf)
http://www.cfsph.iastate.edu/Factsheets/pdfs/boophilus_microplus.pdf; Informe CFSPH, 2007.

De la Fuente J., Rodríguez M., Frago H., Ortiz M., Luiz Massard C., García O., García-García J.C.; Leonart R. (1995). Efficacy of vaccination with Gavac™ in the control of *Boophilus microplus* infestations. In: J. de la Fuente, Editor, *Recombinant Vaccines for the Control of Cattle Tick*, Elfos Scientiae, La Habana, Cuba, 177–185.

Frago H., Hoshman Rad P., Ortiza M., Rodríguez M., Redondo M., Herrera L.; de la Fuente J. (1998). Protection against *Boophilus annulatus* infestations in cattle vaccinated with the *B. Microplus* Bm86-containing vaccine Gavac. *Vaccine*, **16**(20), 1990-1992.

García-García, C. Montero, M. Redondo, M. Vargas, M. Canales, O. Boue, M. Rodríguez, M. Joglar, M. Machado, L. Llina, M. Valdes, L. Mendez; J. De la Fuente. (2000). Control of ticks resistant to immunization with Bm86 in cattle vaccinated with the recombinant antigen Bm95 isolated from the cattle tick, *Boophilus microplus*. *Vaccine*, **18**, 2275–2287.

González, L., Cremata J., Guanche, J., Ramos Y., Triguero A., Cabrera G., Montesino R., Huerta V., Pons T., Boué O., Farnós O. and Rodríguez M.; (2004). The cattle tick antigen, Bm95, expressed in *Pichia pastoris* contains short chains of N- and O-glycans. *Archives of Biochemistry and Biophysics*, **432**, 205-211.

Graphical Codon Usage Analyzer, <http://gcu.schoedl.de/>
Primer BLAST, http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=NcbiHomeAd.

Guglielmone A. y Mangold A. (2007); Garrapata Común de los Bovinos; *idia XXI*; 132-136.

Manuales disponibles en www.stratagene.com de: *E. coli* BL21(DE3).

Manuales disponibles en www.invitrogen.com de: *E. coli* DH5 α , EcoRI, HindIII, NdeI. Buffers Compatibility.

- Martins, J.R.; Correa, B.L.; Cereser, V.H. & Arteché, C.C.P.A. (1995). Situation report on resistance to acaricides by the cattle tick *Boophilus microplus* in the state of Rio Grande do Sul, Southern Brazil. En: “*Memorias III Seminario Internacional de Parasitología Animal*”. (Rodríguez C., S. and Fragosó S., H., ed.). Acapulco. México. 11-13 Octubre: 1-8.
- Menzella H., Ceccarelli E., Gramajo H. (2003). Novel *Escherichia coli* Strain Allows Efficient Recombinant Protein Production Using Lactose as Inducer; *Biotechnology and Bioengineering*, **82**(7), 809-817.
- Nari, A.; Salles, J.; Gil, A.; Waller, P.J. and Hansen, J.W. (1996). The prevalence of anthelmintic resistance in nematodes parasites in sheep in Southern Latin America: Uruguay. *Veterinary Parasitology*, **62**: 213- 222.
- Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO); Resistencia a los Antiparasitarios: Estado actual con énfasis en América Latina; Producción y Sanidad Animal 157; ISSN 1014-1200.
- Panda, A., Khan, R., Rao, A., Totey, S. (1999). Kinetics of inclusion body production in batch and high cell density fed-batch culture of *Escherichia coli* expressing ovine growth hormone. *Journal of Biotechnology*, **75**, 161-172.
- Rodríguez M., Massard C., Henrique da Fonseca A., Fonseca R.N., Machado H., Labarta V.; de la Fuente J. (1995). Effect of vaccination with a recombinant Bm86 antigen preparation on natural infestations of *Boophilus microplus* in grazing dairy and beef pure and cross-bred cattle in Brazil. *Vaccine* **13**, 1804–1808.
- Romero, A.; Benavides, E.; Herrera, C.; Parra, M.H. (1997). Resistencia de la garrapata *Boophilus microplus* a acaricidas organofosforados y piretroides sintéticos en el departamento del Huila. *Revista Colombiana de Entomología*, **23**(1-2), 9-17.
- Thiry M., Cingolani D. (2002). Optimizing scale-up fermentation processes; *Trends in Biotechnology*; **20**(3),103-106.
- Willadsen, P., Bird, P., Cobon, G.S. ; Hungerford, J., (1995). Commercialization of a recombinant vaccine against *Boophilus microplus*. *Parasitology* **110**, S43–S50.
- Willadsen, P., Smith, D., Cobon, G.; McKenna, R.V. (1996). Comparative vaccination of cattle against *Boophilus microplus* with recombinant antigen Bm86 alone or in combination with recombinant Bm91. *Parasite Immunol.* **18**, 241–246.
- Willadsen, P., Kemp, D.H., (1999). Past, present and future of vaccination against ticks. In: Proceedings of the IV Seminario Internacional de Parasitología Animal, Puerto Vallarta, Mexico, October 20–22, 131–140.
- Willadsen, P. (2001). The molecular revolution in the development of vaccines against ectoparasites. *Veterinary Parasitology*, **101**, 353–367.
- Willadsen, P. (2004). Anti-tick vaccines. *Parasitology*, **129**, S367–S387.

ella: – Tío... ¿vos sabés por qué el cielo es celeste?
yo: – Resulta que la luz que viene del sol choca con...
ella: (suspirando) – Tío... ¿viste que lindo que es el cielo?

Diálogos con Victoria

Capítulo 5
***Optimización Experimental de Producción de rBm95 utilizando
Diseño de Experimentos Dinámicos basado en
Modelos de Tendencias Probabilísticos***

5.1. Hipótesis de Trabajo

Es posible utilizar el Diseño de Experimentos Dinámicos basado en Modelos de Tendencias Probabilísticos para la optimización de la producción de rBm95 en cultivos de alta densidad celular de *Escherichia coli* modificada genéticamente.

5.1. *Escherichia coli* en cultivos de alta densidad celular.

5.1.1. Consideraciones generales.

Escherichia coli es quizás el sistema de expresión más utilizado para la producción de proteínas heterólogas. Ese es el microorganismo procariota mejor caracterizado respecto a su genética, fisiología y sistema de expresión de proteínas (Choi, 2000; Lee, 2004; Makrides, 1996). Aunque *E. coli* no puede ser utilizada para producir proteínas de alta masa molecular y/o complejas o que requieren modificaciones post-traduccionales para presentar actividad biológica, este microorganismo sí es adecuado para la obtención de proteínas como interferones, interleukinas, factores estimulantes de colonias, hormonas de crecimiento, sólo por mencionar las más importantes (Lee S, 1996, Gerigk, 2002; Jeong, Lee, 1999; Jeong, 2004; Mijts y Schmidt-Dannert, 2003). Muchas de esas son obtenidas en forma insoluble como cuerpos de inclusión y por lo tanto inactivas requiriéndose pasos de desnaturalización y replegamiento para devolverle su actividad (Kane y Hartley, 1988).

En las primeras etapas de la producción de proteínas recombinantes la investigación se centró en técnicas basadas en la productividad (cantidad de producto por unidad de volumen y de tiempo) ya que las empresas producían biofármacos en una escala pequeña obteniendo buenas ganancias gracias a los altos precios a los que vendían su producción (Datar, 1993). Sin embargo, cuando en el mercado existen diferentes oferentes esa visión necesariamente cambia y la reducción en el costo de producción se vuelve un factor más importante.

Debido a que la mayoría de las proteínas producidas en *E. coli* son acumuladas intracelularmente, la productividad es proporcional a la densidad celular final y a la productividad específica (cantidad de producto por unidad de masa celular y tiempo). Por lo anterior, han sido desarrolladas técnicas de cultivos de alta densidad celular con el objetivo de mejorar la productividad y además concomitantemente obtener otras ventajas como la reducción del volumen de los cultivos, mejorar los procesos

downstream, reducir los efluentes, bajar la inversión en equipamiento, bajar los costos de producción, etc (Choi, 2006).

La estrategia *fed-batch* es la más utilizada para obtener cultivos de alta densidad celular (Kleman y Strohl, 1994; Yee y Blanch, 1992; Riesenber, 1991). El objetivo de esas técnicas es lograr el crecimiento sostenido del microorganismo para lo cual son adoptadas diferentes estrategias y vehículos para la alimentación de nutrientes al biorreactor. Sin embargo, dependiendo de la técnica se presentan dificultades como inhibición por sustratos (Kazan, 1995), la capacidad limitada para la transferencia de oxígeno a la fase líquida (Bhattacharya y Dubey, 1997), la formación de sub-productos tóxicos, entre otras (Roe, 2002; Luli y Strohl, 1990). Otra complicación, que ha sido demostrada en los cultivos de alta densidad celular, es que la productividad específica de las proteínas recombinantes generalmente es menor que la obtenida en los cultivos a escala laboratorio (Jeong y Lee, 1999).

Existen diferentes aproximaciones teóricas de la concentración celular máxima alcanzable en un reactor. Riesenber (Riesenber, 1991) en una primera aproximación calculó para una condición de reactor empacado con células un valor máximo de 400 gDCW.L⁻¹. Más adelante, Markl y col. (Markl, 1993) utilizaron un modelo celular correspondiente a un bacilo de 3 µm de largo por 1 µm de diámetro y concluyeron que para un reactor empacado como mínimo el 25% del volumen debía ser ocupado con medio de cultivo. Si el peso seco celular es entre un 20-25% del peso húmedo y considerando que la densidad de las células es poco mayor que la del agua, entonces la máxima concentración celular alcanzable es de 160-200 gDCW.L⁻¹. Estos valores parecen razonables a juzgar por las densidades celulares alcanzadas: 174 gDCW.L⁻¹ para *E. coli* W3110 en un reactor con diálisis (Markl, 1993) y 175.4 gDCW.L⁻¹ para *E. coli* recombinante productora de poly(3-hydroxybutyrate) (Lee y Chang, 1994).

Por otro lado, la viscosidad del medio de cultivo crece dramáticamente cuando la concentración celular supera los 200 gDCW.L⁻¹ y el caldo pierde su fluidez cuando la densidad celular alcanza los 220 gDCW.L⁻¹ (Mori, 1979).

Otra característica importante que diferencia a los sistemas biológicos de aquellos en los que ocurren reacciones químicas es que algunos nutrientes pueden producir la inhibición del crecimiento celular si se encuentran en el medio de cultivo por encima de un valor umbral (Riesenber, 1991). La Tabla 5.1 contiene valores de concentraciones máximas pre-inhedoras de algunos nutrientes.

Tabla 5. 1: Concentraciones máximas pre-inhedoras de nutrientes en cultivos de *E. coli*

Nutriente	Valor límite (g/L)
Glucosa	50
Amonio	3
Hierro	1,15
Magnesio	8,7
Fósforo	10
Zinc	0,038

Además, existen trabajos en los que ha quedado demostrado que la presencia de glucosa en concentraciones superiores a 1,5-2 g.L⁻¹ produce un desacople metabólico cuya consecuencia es la inhibición del crecimiento celular producto de la acumulación de acetato (Doelle y col., 1974; Luli y Strohl, 1990) (ver más adelante).

Shiloach y Fass (Shiloach y Fass, 2005) reportan que en un medio definido con los valores máximos no inhibitorios de nutrientes permiten alcanzar una producción celular de 15 gDCW.L⁻¹.

El acetato es el principal subproducto en los cultivos de *E. coli* y es generado en dos situaciones:

- Cuando el cultivo es anaerobio o presenta limitaciones en la concentración de oxígeno disuelto.
- Cuando el flujo de carbono en la vía metabólica central excede la demanda biosintética y la capacidad de generación de energía en las células. Eso produce un desbalance metabólico generado por la saturación en el ciclo del ácido tricarbóxico y/o en la cadena de transporte de electrones conocido como efecto Crabtree (Luli y Strohl, 1990).

Este último efecto se ve particularmente favorecido en las condiciones de cultivos *fed-batch* en los que existen altas concentraciones de glucosa y una alta velocidad de crecimiento celular.

Una alta concentración de acetato (por encima de los 5 g.L⁻¹ a pH 7) reduce la velocidad de crecimiento, el rendimiento en biomasa y la densidad celular máxima alcanzable independientemente de si la cepa de *E. coli* utilizada puede o no metabolizar dicho anión o si ha sido modificada genéticamente (Luli y Strohl, 1990; E1-Mansi y Holms1989; Kleman y Strohl, 1994; Holms, 1986; Varma y Palsson, 1994; Pan y col., 1987; Han y col., 1992; Jensen y col., 1990). Es por eso que existe un amplio consenso

en que la técnica de propagación utilizada debe asegurar una concentración de acetato por debajo de 2 g.L^{-1} (Shiloach, 2005).

La forma protonada del acetato extracelular puede atravesar la membrana celular debido a su carácter lipofílico y, como el citoplasma tiene un pH mayor al del medio (7,5 y controlado en 7, respectivamente), se produce un ingreso constante de esa especie produciendo una disminución del pH intracelular. Esto interfiere en la producción de ATP ya que reduce el ΔpH a través de la membrana celular y por lo tanto en la fuerza protón-motriz además de consumir mayor cantidad de energía para el mantenimiento del equilibrio homeostático intracelular.

Por otro lado, el acetato tiene efectos negativos en la producción de proteínas tanto en células recombinantes como no recombinantes (Shimizu, 1992; MacDonald y Neway, 1990).

La formación de acetato depende del medio utilizado y la velocidad de crecimiento específica ya que es formado en medios tanto indefinidos como definidos cuando la velocidad de crecimiento específica supera los $0,2$ o $0,35 \text{ h}^{-1}$, respectivamente (Paahne, 1990; El-Mansi y Holms, 1989; Meyer, 1984). Además, de un estudio con diferentes cepas de *E. coli* ha sido reportado que hubo producción de acetato en un medio definido a velocidades de crecimiento específico tan bajas como $0,14 \text{ h}^{-1}$ (Fieschko, 1989). Por lo tanto, la velocidad de crecimiento específica que lleva a la producción de acetato varía según la cepa y la composición del medio de cultivo utilizado.

La estrategia de alimentación del biorreactor es crítica para la obtención de cultivos de alta densidad celular. Dicha estrategia no sólo afecta a la concentración celular máxima alcanzada sino también a la productividad celular ya que la formación de la proteína recombinante puede variar dependiendo de la estrategia de alimentación implementada (Jung, 1988; Jensen, 1990).

Los cultivos de alta densidad celular son realizados en condiciones de limitación de nutrientes (generalmente la fuente de carbono) pero también se han logrado altos rendimientos en biomasa implementando políticas de alimentación constantes (Markl, 1993; Jensen, 1990; Jung, 1988), en escalón (Konstantinov, 1991; Paahne, 1990) o con un caudal que crece exponencialmente (Pan, 1987; Yoon, 1994; Babu, 2000; Qiu, 1998; Sletta, 2004; Wong, 1998).

Cuando una política de alimentación constante es utilizada la velocidad de crecimiento específica disminuye continuamente debido al aumento tanto del volumen del cultivo como de la concentración celular. La alimentación en escalón o gradual puede aumentar

la velocidad de crecimiento ya que agrega una mayor cantidad de nutrientes a concentraciones celulares más altas. Por último, si se adopta una política de alimentación del tipo exponencial puede lograrse que el microorganismo crezca exponencialmente durante todo cultivo y además puede disminuirse la producción de acetato si se mantiene la velocidad de crecimiento específica por debajo del valor crítico para la formación de acetato ($0,1-0,3 \text{ h}^{-1}$) (Strandberg y Enfors, 1991; Lee y Chang, 1993; Helmuth, 1994).

Existen otras estrategias de alimentación de biorreactores más sofisticadas que están basadas en esquemas de control por retroalimentación en los que el caudal de alimentación se encuentra acoplado a los valores de parámetros físicos medidos *on-line* como DO (Mori, 1979; Eppstein, 1989; Seo, 1992), pH (Lee y Chang, 1993; Lee y Chang, 1994; Mizutani, 1986), la evolución de cambio de CO_2 (CER) (Paahne, 1990; Chao, 2002; Hu, 2004; Kleist, 2003; Shin, 1997; Yuan, 2004).

Hasta el momento no hay una regla general que permita seleccionar una estrategia de alimentación adecuada que garantice ni la máxima producción celular ni la mayor productividad de proteína heteróloga. Es por eso que suelen ensayarse diferentes políticas de alimentación en un esquema de prueba y error que suele no ser sencillo (Choi, 2006). Sin embargo, las mayores productividades en los cultivos de alta densidad celular han sido alcanzadas cuando las fases de crecimiento y producción están separadas. En los cultivos *fed-batch* esto se logra retrasando la inducción hasta que el cultivo haya alcanzado una concentración celular adecuada. Por lo tanto, la determinación de las condiciones de inducción óptimas es importante para aumentar la productividad global de la proteína recombinante (Choi, 2000; Jeong y Lee, 1999; Yim, 2001).

5.1.2. Scaling up:

El cambio de escala de un proceso es complicado en general pero particularmente lo es cuando son considerados bioprocesos ya que en estos casos debe tenerse en cuenta, y por lo tanto haber caracterizado anteriormente, todo el ambiente en el que se encuentra una célula (Young, 1979). Entre los factores biológicos que se ven afectados por la escala se encuentra el número de generaciones asociado con el desarrollo del inóculo y la fase de producción, la probabilidad de mutación, la vulnerabilidad a la contaminación, la formación de *pellets* y la presión selectiva (Enfors, 2001; Ettler, 1990; Sweere, 1987; Trilli, 1986). Entre los factores químicos que se ven afectados por

la escala pueden enumerarse los agentes químicos utilizados para controlar el pH (tanto el tipo como la concentración), la calidad de los componentes del medio de cultivo, la calidad del agua, por mencionar algunos. Un factor muy importante es la concentración de oxígeno disuelto en el medio de cultivo ya que ha sido demostrado que influye tanto en la producción celular como la expresión de las proteínas recombinantes (Bhattacharya y Dubey, 1997; Horn, 1996).

5.2. Descripción del biorreactor

- El biorreactor utilizado para realizar los cultivos de *E. coli* BL21 transformada con la construcción pET-24b-*bm95* fue el fermentador modular escala banco BioFlo 110 comercializado por la empresa New Brunswick Scientific Co., Inc. La Figura 5.1 muestra una ilustración gráfica, y la Figura 5.2 una vista fotográfica del sistema BioFlo 110 montado en el laboratorio.

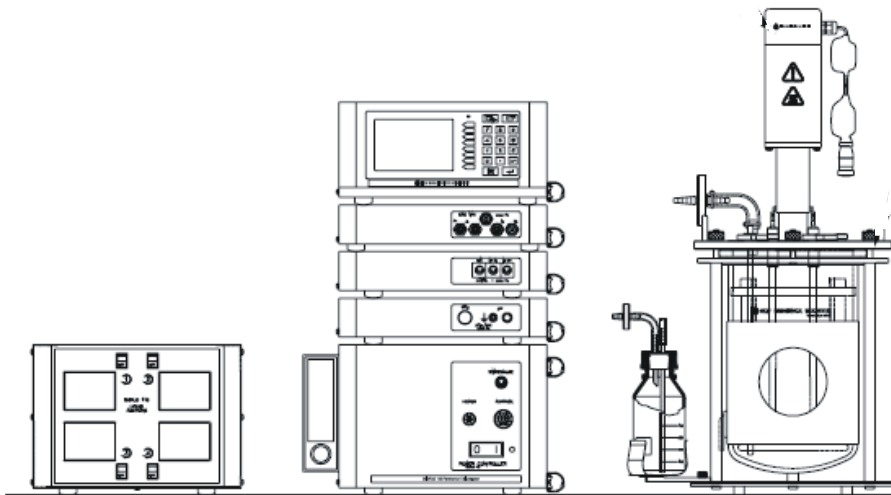


Figura 5.1: Ilustración del sistema BioFlo 110.

El vaso del biorreactor es de vidrio y está graduado para estimar el volumen del líquido en su interior. Tiene una capacidad máxima de 3 L., con un volumen de trabajo mínimo de 800 mL (definido por la altura a la que los sensores de pH y pO₂ quedan sumergidos) y uno máximo recomendado de 2,2 L.

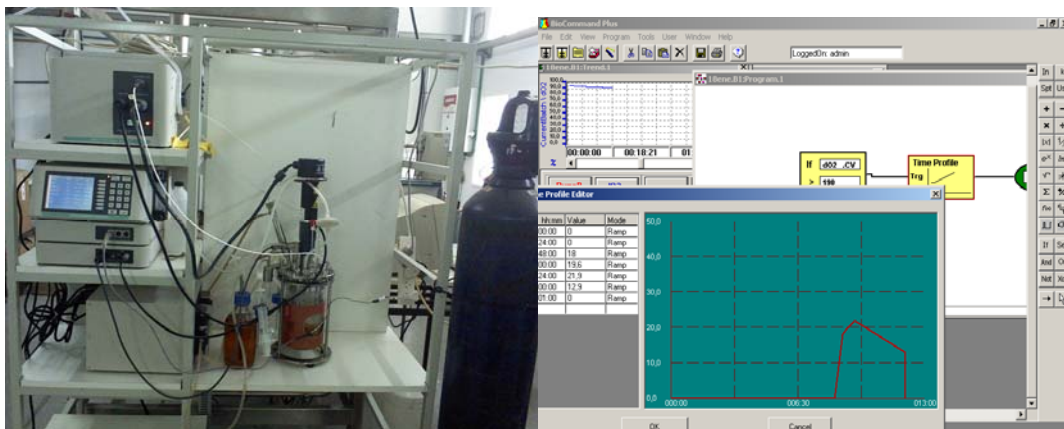
El equipo cuenta con los siguientes módulos independientes:

- Unidad de control primaria.
- Controlador de pH y pO₂.

- Controlador de nivel.
- Controlador de temperatura y agitación.
- Módulo de 4 bombas peristálticas.

Además, el sistema cuenta con:

- Rotámetro.
- Camisa calefactora.
- Válvula solenoide para ingreso de agua de refrigeración.
- Serpentina para refrigeración.
- Motor Agitador con velocidad de rotación variable.
- 4 baffles desmontables.
- Dos agitadores de 6 paletas planas de posición vertical variable sobre el eje.
- Vaso graduado de 3 L de capacidad.
- Refrigerante para gases de salida.
- Filtros de esterilización de gases de entrada y salida.
- Sparger.
- Soporte de acero inoxidable para montaje del vaso.
- Sensor de pH.
- Sensor de pO₂ con electrodo polarográfico.
- Dispositivo de muestreo.
- 2 Sensores de nivel.
- Termocupla con termorresistencia de platino.
- Tubo para cosecha.



a

b

Figura 5.2: Vista del sistema BioFlo 110 montado en laboratorio (a) y captura de pantalla del *software* para el monitoreo y control.

Los sistemas de control son descritos en la tabla 5.2.

Tabla 5. 2: Sistemas de control del biorreactor BioFlo 110

Variable controlada	Loop	Controlador	Actuador	Rango de control
Temperatura	Cerrado	PI	Válvula solenoide para ingreso agua refrigeración Camisa calefactora	$\pm 0,1$ °C
pO ₂	Cerrado	PID	Motor de agitación	$\pm 1\%$
pH	Cerrado	PID	Bombas peristálticas con ácido o base	0,2
Caudal alimentación	Abierto/Cerrado		Bombas peristálticas	0,02 mL.min ⁻¹

5.3. Definición de las condiciones de cultivo:

5.3.1. Parámetros constantes de la política de operación.

De los resultados obtenidos a en los ensayos de expresión a escala laboratorio (cultivos de 5 mL) el medio de cultivo escogido para las experiencias a escala banco (volumen final 1,8 L) fue TB suplementado con glucosa debido a la mayor masa celular y proporción de rBm95 obtenida en dichos experimentos (ver capítulo 4, sección 4.3.10). El cultivo fue llevado a cabo en medio no selectivo (kan[-]) en base a los siguientes ensayos de estabilidad:

- determinada en ensayos a escala laboratorio (capítulo 4, sección 4.3.10): 95% a las 8 hs de cultivo.
- evaluada a escala banco con un cultivo de 1L al que se le adicionó kanamicina hasta alcanzar una concentración de 30 mg.L⁻¹ al momento de la inducción, determinado la biomasa al momento de la inducción y 4 hs post inducción no encontrándose una diferencia significativa: $9,3 \pm 0,5$ g.L⁻¹ y $8,9 \pm 0,6$ g.L⁻¹ ($1 - \alpha = 0.95$), respectivamente.

La concentración de oxígeno disuelto se mantuvo controlada por encima del valor mínimo del 20% de saturación en la fase líquida utilizando la velocidad de agitación como variable manipulada. El aporte de dicho nutriente se realizó según lo sugerido por Shiloach y Fass (Sholoach y Fass, 2005) mediante la inyección de oxígeno puro a un caudal constante de 1 LPM (Bhattacharya y Dubey, 1997; Luli y Strohl, 1990).

La ventana temporal a ajustar por los modelos de tendencia y de la cual han sido seleccionados los datos experimentales para estimar los parámetros de los modelos de la biblioteca ha sido definida en base a experiencias exploratorias. Así, el primer punto

considerado en la estimación no corresponde con el inicio de la corrida sino que con el inmediatamente posterior al cambio de pendiente en la caída de la concentración de oxígeno disuelto, lo cual es indicativo del comienzo del crecimiento exponencial de *E. coli* (fin de la fase lag), como está indicado en la Figura 5. 3.

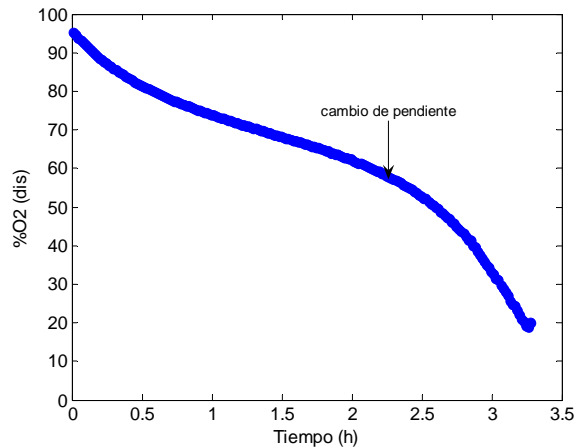


Figura 5.3: Dinámica de oxígeno para cultivo representativo de *E. coli* BL21 transformada con la construcción pET-24b-*bm95*

El tiempo final que se incluyó en la estrategia de optimización no se corresponde con el final de la corrida sino con el tiempo al cual el cultivo es inducido. Esto último se debe a que ha sido corroborado experimentalmente lo descrito por Bhattacharya y Dubey (Bhattacharya y Dubey, 1997) quienes reportan que el agregado de IPTG al medio de cultivo produce un cambio inmediato en la velocidad de crecimiento celular similar a una fase lag durante la cual la proteína heteróloga es expresada en detrimento de las proteínas celulares necesarias para el crecimiento y el metabolismo. La Figura 5.4 muestra ese fenómeno. Además, esto es conveniente debido a que ha sido comprobado que la cepa transformada produce mejores rendimientos de rBm95 cuando es inducida a 30°C. El contenido de proteínas totales de *E. coli* BL21 transformada con la construcción pET-24b-*bm95* en esas condiciones fue del $[20,6 \pm 2,4]\%$ del residuo seco. Del análisis cuantitativo de la imagen de un SDS-PAGE en el que se corrió una muestra del lisado total se determinó que la banda de sobreexpresión correspondía a un $(2,57 \pm 0,3)\%$ del DCW a las 4 hs de inducción.

Por lo tanto, la fase de crecimiento celular del bioproceso a optimizar se realizó en condiciones isotérmicas a 37 °C.

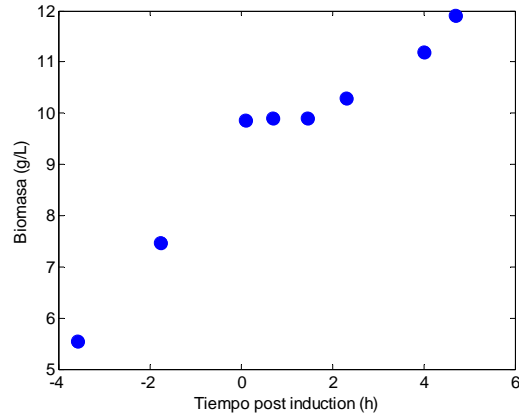


Figura 5.4: Cambio metabólico en *E. coli* BL21 transformada con la construcción pET-24b-*bm95* debido a la inducción.

La Tabla 5.3 resume de las condiciones de cultivo que se han mantenido constantes en todas las experiencias de optimización que fueron realizadas.

Tabla 5. 3: Parámetros de la política de operación fijados *a priori* de la optimización

Parámetro	Valor fijado
Límite inferior para $O_{2(\text{dis})}$	20 % saturación
Rango de velocidad de agitación	[200-600] rpm
Temperatura fase de crecimiento	37 °C
Temperatura fase inducción	30 °C
Tiempo inducción	4 hs
pH	7,0
Antiespumante	Antifoam 204 [0,1% (v/v)]

5.3.2. Parámetros manipulables de la política de operación

El rango de variación permitido para la concentración de sustrato inicial ha sido definido entre 5 y 14 g.L^{-1} , el valor límite superior está un poco por debajo de lo reportado por Kazan (Kazan, 1995) de manera de evitar la inhibición del crecimiento celular debido a altas concentraciones de glucosa.

Debido a que en cada experimento para optimización se busca mantener un control estricto de la política de alimentación de sustrato, definida por la metodología de optimización basada en modelos de tendencia, se definió una concentración máxima de glucosa en la alimentación de 45 g.L^{-1} lo que permitió operar caudales de alimentación superiores al 5% del caudal máximo de la bomba peristáltica del biorreactor BioFlo 110.

La política de alimentación del biorreactor adoptada para la optimización corresponde a un polinomio inverso de segundo orden como el descrito en la ec. 5.1. Este tipo de formulación ha demostrado tener una gran versatilidad en los perfiles que pueden obtenerse conservando un bajo número de coeficientes (Martinez, 2009).

$$F_{in} = \begin{cases} 0, & t < t_{feed} \\ \frac{At'}{1 + Bt' + (Ct')^2}; t' = (t - t_{feed}), & t \geq t_{feed} \end{cases} \quad (5.1)$$

Para definir los intervalos de variación permitidos para los parámetros de la política de alimentación A , B , C y t_{feed} ha sido realizado un estudio de simulación restringiendo el caudal mínimo a 0 L.h⁻¹ (condición de no descarga) y el máximo en 1,3 L.h⁻¹ (caudal máximo de la bomba).

En experiencias exploratorias fue observado que el cultivo de *E. coli* BL21 transformada con la construcción pET-24b-*bm95* con concentraciones iniciales de glucosa (S_0) entre 10 y 14 g.L⁻¹, y de biomasa entre 0,018 y 0,050 g.L⁻¹ consumían completamente la fuente de carbono entre las 7 y 8,5 hs de cultivo por lo tanto el tiempo de inicio de la alimentación (t_{feed}) se definió en el intervalo [4-10] hs. Además, debido a que no se deseaba que las experiencias (fase de crecimiento + fase inducción) se prolongasen más de 18 hs, el tiempo máximo tanto de cultivo (t_{final}) como de finalización de la alimentación ($t_{feedend}$) se fijó en 14 hs. En la tabla 5.4 es mostrado un resumen del espacio W definido para la búsqueda del óptimo.

Tabla 5. 4: Valores límites de los parámetros manipulables de la política de operación

Parámetro	Unidades	Valores	
		inferior	superior
A	L.h ⁻²	0,0136	0,9981
B	h ⁻¹	0,0031	0,9915
C	h ⁻²	0,0038	0,9961
t_{final}	h	10	14
t_{feed}	h	4	10
$t_{feedend}$	h	10	14
S_0	g.L ⁻¹	5	14
S_{feed}	g.L ⁻¹	30	40

5.4. Biblioteca de Modelos de Tendencias para Optimización.

La biblioteca de modelos de tendencias para optimización fue construida con 4 modelos: Monod, Contois, Mantenimiento y 2° orden. Tal como se definió a estos modelos en el capítulo II del presente trabajo, ellos tienen estructuras simples y bajo número de parámetros (4 para Monod y Contois y 5 para Mantenimiento y 2° orden). Con ellos se intentó describir la dinámica de las variables de estado biomasa (X) y

glucosa (S) medidas off-line según es descrito en el Anexo 5. Las ecuaciones 5.2 muestran la estructura general de los modelos y las 5.3-6 las características estructurales particulares para cada modelo de la biblioteca (para simplificar la notación se eliminaron las dependencias temporales de las variables y los símbolos siguen la nomenclatura seguida en el capítulo II).

$$\begin{aligned} \frac{dV}{dt} &= F; \quad D = \frac{F}{V} \\ \frac{dX}{dt} &= \mu X - DX; \\ \frac{dS}{dt} &= -\frac{\mu}{Y_{XS}} X - f(X, t) + D(S_{feed} - S); \end{aligned} \quad (5.2)$$

Monod $\mu = \mu_{\max} \frac{S}{K_S + S}; \quad f(X, t) = 0$ (5.3)

Contois $\mu = \mu_{\max} \frac{S}{K_S X + S}; \quad f(X, t) = 0$ (5.4)

Mantenimiento $\mu = \mu_{\max} \frac{S}{K_S + S}; \quad f(X, t) = mX$ (5.5)

2º orden $\mu = \mu_{\max} \frac{S}{K_S + S + \left(\frac{S}{KI}\right)^2}; \quad f(X, t) = 0$ (5.6)

5.5. Mejora de la producción de rBm95

Siguiendo la tendencia marcada por Choi (2006) la función escogida para optimización ha sido la producción de *E. coli* BL21 transformada con la construcción pET-24b-*bm95* (ec. 5.7).

$$J = X(t_{final}) - V(t_{final}) \quad (5.7)$$

El Método de Diseño de Experimentos Dinámicos para Optimización basado en Modelo de Tendencias presentado en el capítulo II ha sido implementado para la mejora de la función objetivo (ec. 5.7).

Las políticas de operación del bioproceso tanto las calculadas por el método w^k , $k =$ Monod, Contois, ...; como las implementadas en la realidad w^{real} (debido a problemas técnicos y/o errores experimentales presentes en todo ensayo) durante la optimización son mostradas en la tabla 5.5. La política de operación en cada iteración fue calculada

con un modelo de tendencias seleccionado en base al error total E_{total} . Además, en la tabla 5.5 es mostrado el valor de la función objetivo J obtenido en forma experimental. En la figura 5.5 son mostrados los perfiles de alimentación del biorreactor implementados en cada experiencia durante la optimización.

Las estrategias de muestreo óptimas para cada experimento óptimo calculadas con las restricciones: $0.5h \leq t_{i+1}^{muestra} - t_i^{muestra} \leq 3h$ y $n_{sp} = 10$, fueron:

$$inicial \left\{ \psi_0 = [2.5, 5, 6, 7, 8, 9, 9.5, 10, 10.5, 11, 11.5, 12] \right.$$

$$1^\circ \text{ iter} \left\{ \begin{array}{l} \psi^{opt}(h) = 0, 1.16, 3, 4.95, 6.05, 7.8, 8.41, 10.22, 12.09, 13.45 \\ M = 9 \times 10^{-30} \end{array} \right.$$

$$2^\circ \text{ iter} \left\{ \begin{array}{l} \psi^{opt}(h) = 2.72, 5.49, 6.33, 8.11, 8.82, 9.55, 11.32, 12.38, 13.03, 13.9 \\ M = 2.44 \times 10^{-24} \end{array} \right.$$

$$3^\circ \text{ iter} \left\{ \begin{array}{l} \psi^{opt} = 3.6, 4.72, 6.09, 7.57, 8.43, 9.24, 9.86, 10.57, 11.44, 12.9 \\ M = 1.25 \times 10^{-18} \end{array} \right.$$

Un rasgo característico es que en todas las experiencias la metodología propone que los muestreos se hagan más espaciados al principio de las corridas, cuando los cambios no muy significativos, y que la frecuencia aumente en los momentos inmediatamente anteriores y posteriores al comienzo de la alimentación.

Tabla 5.5: Políticas de operación calculadas (cursivas) e implementadas durante la optimización

Parámetro	Experiencia							
	w_0^{real}	$w_1^{2^\circ \text{orden}}$	w_1^{real}	$w_2^{Contois}$	w_2^{real}	$w_3^{Contois}$	w_3^{real}	$w_4^{2^\circ \text{orden}}$
F	0,198							-
A		0,0956	0,0956	0,8480	0,8480	0,8330	0,8330	0,7588
B		0,0440	0,0440	0,9915	0,9915	0,4530	0,4530	0,7608
C		0,0867	0,0867	0,9961	0,9961	0,9209	0,9209	0,9687
t_{final}	9	12	12	14	14	12,9	12,9	13,7
t_{feed}	6	7,5	7,5	8,71	9,3	9,30	9,4	9,30
$t_{feedend}$	8,5	12	11,8	12	12,3	12	12	12
S_0	9,25	10	10,71	12	14	12,86	13,34	12,9
S_{feed}	28,15	35	39	38	40	35	33,12	37,5
J	10,8		16,0		20,3		17,0	

La aplicación del método de optimización para la mejora iterativa de la producción de *E. coli* BL21 transformada con la construcción pET-24b-bm95 comenzó con una experiencia conservadora (Experiencia 0), tanto desde el punto de vista de la política de operación w_0^{real} , en la que el perfil de carga presenta la forma de escalón, como de la estrategia de muestreo ψ_0 .

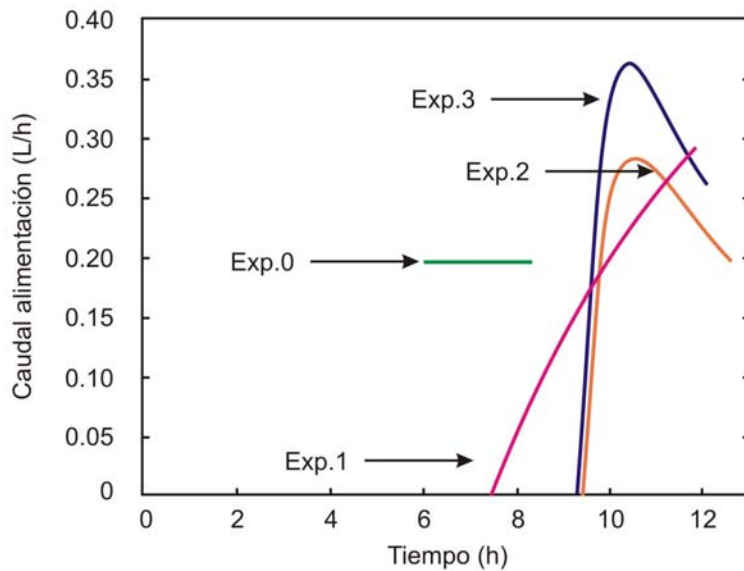


Figura 5.5: Políticas de alimentación del biorreactor BioFlo 110 durante el proceso de optimización

Considerando los errores experimentales inherentes a la implementación de una política de operación del biorreactor, el método converge luego del diseño de 4 experiencias para optimización a una política en la que son obtenidos 20,3 gDCW lo que significaría una producción estimada de aproximadamente 522 mg de rBm95 (equivalente a 251 dosis de una formulación vacunal). Los parámetros de $w_4^{2^{\circ}orden}$ que no están relacionados con la alimentación son indistinguibles de los de w_2^{real} desde el punto de vista experimental. Sin embargo, si bien los coeficientes que definen F_{in} en $w_4^{2^{\circ}orden}$ no convergen a los valores obtenidos en w_2^{real} , al implementar los caudales se obtienen los mismos perfiles temporales (figura 5.6).

Aunque el valor optimizado está lejos de alcanzar los reportados en la bibliografía para biorreactores *fed-batch* sin diálisis (Lee, 1996; Choi, 2006), cabe señalar que para alcanzar esas concentraciones celulares los tiempos de cultivos fueron mucho mayores. Además, al comparar nuestros resultados con trabajos en los que son mostradas las evoluciones de la concentración celular (Korz, 1995; Luli y Strohl, 1990; Panda, 1999), estos resultan comparables para tiempos de cultivos similares a los implementados en este trabajo de tesis.

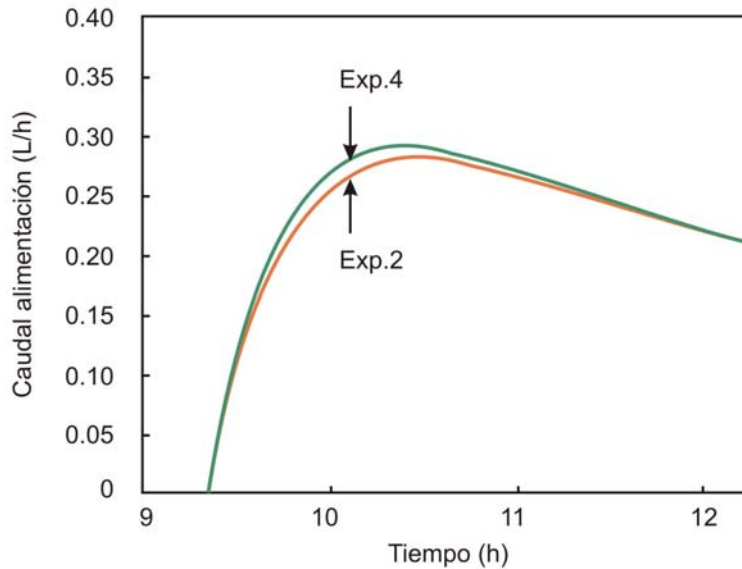


Figura 5.6: Comparación de Políticas de alimentación diseñadas en la experiencia 2 y la considerada de convergencia

En la figura 5.7 son mostradas las trayectorias de la función objetivo medidas experimentalmente durante el proceso de optimización junto con las estimaciones realizadas con los modelos de tendencias. En esa figura puede ser observado el aumento de la distancia entre las predicciones de la biblioteca de modelos y la trayectoria real de J a medida que el proceso es mejorado. Ese efecto es cuantificado en la tabla 5.6 mediante el cálculo de los errores paramétrico $E_{j_i \rightarrow \{j\}}$, estructural $E_{\langle j \rangle \rightarrow j}$ y total E_{total} .

La pérdida de calidad en las estimaciones de la biblioteca de modelos de tendencias puede deberse al aumento en la rigidez de los modelos, producida principalmente por el estrechamiento de los espacios de incertidumbre paramétricos (ver más adelante). La consecuencia es la pérdida de calidad para la extrapolación hacia zonas de operación diferentes del proceso. Esta característica ya había sido observada en los ejemplos computacionales de optimización en los que los modelos presentaban diferencias estructurales con el biorreactor *in silico* (Capítulo 2, sección 2.4 y Capítulo 3, sección 3.4) a optimizar.

En las tablas 5.7 y 5.8 son mostrados los perfiles de sensibilidad de las estimaciones de los modelos de tendencias durante la optimización experimental luego de realizar un GSA (en negrita aparecen los parámetros re-estimados en cada iteración). La sensibilidad de los modelos varía dependiendo de la condición de operación en la que el GSA es realizado. Sin embargo, puede observarse que en los modelos existen

parámetros que debieron ser re-estimados más de una vez (μ_{max} para Monod, k_{lis} para Contois, m para Mantenimiento y Y_{xs} para 2° orden) mientras que otros no fueron re-estimados nunca (k_{lis} para Mantenimiento y Ks para 2° orden).

Con los datos recolectados en la experiencia 0 (w_0^{real}, ψ_0) fueron definidos los espacios de incertidumbre paramétrica inicial Θ_0^k ; $k = \text{Monod, Contois, ...}$. En la Tabla 5.9 son mostrados los intervalos de confianza para el 95 percentilo de los parámetros de los modelos de tendencias y en las figuras 5.8 y 9 las *fdp* empíricas calculadas utilizando *bootstrap* para los modelos utilizados en la optimización.

Tabla 5. 6: Errores paramétricos, estructurales y totales

Modelo	Experiencia											
	0			1			2			3		
	$E_{j_i \rightarrow \langle j \rangle}$	$E_{\langle j \rangle \rightarrow j}$	E_{total}	$E_{j_i \rightarrow \langle j \rangle}$	$E_{\langle j \rangle \rightarrow j}$	E_{total}	$E_{j_i \rightarrow \langle j \rangle}$	$E_{\langle j \rangle \rightarrow j}$	E_{total}	$E_{j_i \rightarrow \langle j \rangle}$	$E_{\langle j \rangle \rightarrow j}$	E_{total}
Monod	0,1222	1,0445	1,1667	0,2983	7,6038	7,9021	0,0446	14,857	14,901	0,1244	20,322	20,446
Contois	0,0469	2,0650	2,1119	0,1825	7,5000	7,6825	0,0414	8,3727	8,4141	0,0634	51,913	51,973
Mantenimiento	0,0278	1,2275	1,2553	0,5356	16,373	16,909	0,0528	8,5014	8,5542	0,0587	34,467	34,526
2° orden	0,0177	1,1405	1,1582	0,2843	9,2472	9,5315	0,1100	8,9271	9,0371	0,1294	17,744	17,874

Tabla 5. 7: Indices de sensibilidad global de primer orden (experiencias 0 y 1)

Parámetro	Experiencia							
	0				1			
	Monod	Contois	Manten.	2° orden	Monod	Contois	Manten.	2° orden
μ_{max}	0,3407	0,0161	0,5298	0,3037	0,3234	0,0398	0,2994	0,2920
K_S	0,0760	0,0111	0,2914	0,1901	0,0374	0,0377	0,0953	0,0761
Y_{xs}	0,1605	0,3565	0,0648	0,0003	0,1623	0,3534	0,0361	0,2842
K_I				0,0713				0,1365
k_{lis}	0,2366	0,4102	0,0860	0,0931	0,1795	0,3799	0,0179	0,3213
m			0,2275				0,3806	

Tabla 5. 8: Indices de sensibilidad global de primer orden (experiencias 2 y 3)

Parámetro	Experiencia							
	2				3			
	Monod	Contois	Manten.	2° orden	Monod	Contois	Manten.	2° orden
μ_{max}	0,4994	0,7095	0,0060	0,1018	0,0562	0,2622	0,1033	0,0534
K_S	0,2310	0,0686	0,9568	0,0330	0,0540	0,3587	0,0521	0,0210
Y_{xs}	0,1746	0,0732	0,0080	0,0538	0,5320	0,0722	0,3244	0,3210
K_I				0,3523				0,0642
k_{lis}	0,1738	0,1429	0,0018	0,2310	0,0266	0,1359	0,0563	0,4508
m			0,0073				0,4273	

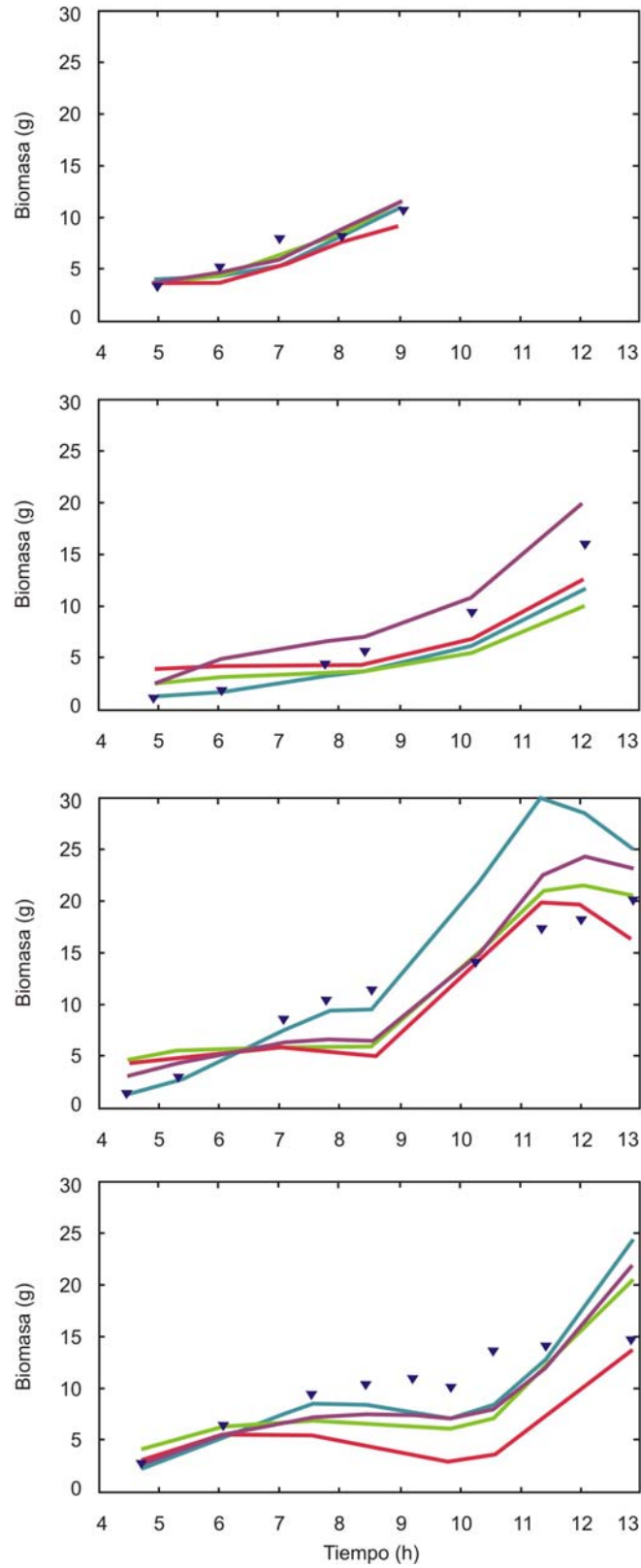


Figura 5.7: Trayectorias de la función objetivo J y las estimaciones de los modelos Monod (verde), Contois (rojo), Mantenimiento (celeste), 2° Orden (violeta) para las experiencias: 0, 1, 2, 3.

Tabla 5. 9: Intervalos de confianza iniciales (95 percentilo) y valor más frecuente de los parámetros de cada modelo de la biblioteca de modelos de tendencias.

Parámetro	Experiencia 0			
	Monod	Contois	Mantenimiento	2º orden
μ_{max}	[0,7005-4,9822] 2,2205	[0,5855-3,4714] 0,8019	[0,7187-3,1992] 2,2938	[1,7729-7,3679] 4,4865
K_s	[1,4457-30,1743] 1,5893	[0,0593-1,1513] 0,3705	[2,9719-31,3276] 11,6204	- [4,9606-51,5902] 29,6743
Y_{xs}	[0,3413- 1,0000] 0,9967	[0,1833-1,0000] 0,9959	[0,6369-1,0000] 0,9982	[0,4777-1,0000] 0,8720
KI	-	-	-	[2,6320-64,4593] 2,9412
k_{lis}	[0,0057- 0,1430] 0,1423	[0,0333-0,7443] 0,2004	[0,0002-0,0055] 0,0055	[0,0060-0,1410] 0,1403
m	-	-	[0,0342- 0,2618] 0,2197	-

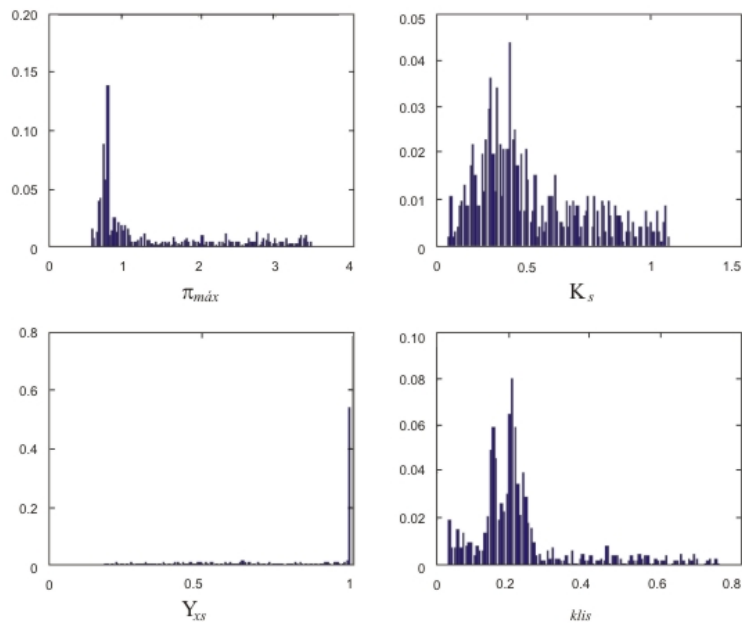


Figura 5.8: *fdp* empíricas iniciales de los parámetros del modelo Contois.

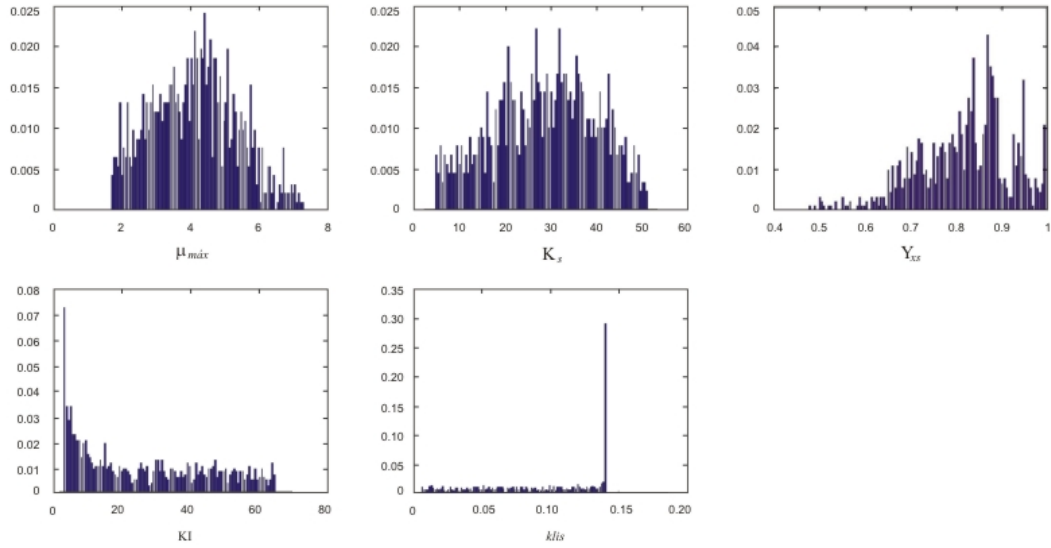


Figura 5.9: *fdp* empíricas iniciales de los parámetros del modelo 2º orden.

Con la implementación de cada política (w_j^{real}, ψ_j) , $j = 1, 2, 3$ los Θ_j^k fueron modificados re-estimando los parámetros más relevantes surgidos del *GSA*. En tablas 5.10 y 5.11 son mostrados tanto los intervalos de confianza para el 95 percentilo como el valor más frecuente de los parámetros de los modelos de tendencias. En las figuras 5.10 y 5.11 son mostradas las evoluciones de las *fdp* de los parámetros re-estimados de los modelos de tendencias utilizados para el Diseño de Experimentos de Dinámicos para Optimización, k_{lis} tanto para el modelo Contois como para el 2º orden. Además, en las figuras 5.12 y 5.13 puede observarse el aumento progresivo de la rigidez de dichos modelos.

En un primer momento, cada modelo presenta una incertidumbre paramétrica Θ amplia permitiendo cubrir un gran espacio de posibles trayectorias estimadas para la función objetivo J . Sin embargo, a medida que los Θ son sesgados, a través de la re-estimación de los parámetros más relevantes, ese espacio de estimaciones posibles de J se estrecha. Esa rigidez se profundiza hasta que en la última experiencia (experiencia 3) los modelos actualizados ya no son capaces de incluir todos los valores tomados realmente por J .

Vale la pena destacar que más allá de la poca capacidad de la biblioteca de modelos de tendencias para ajustar la trayectoria de J , el valor final (que es el considerado para optimización) sigue siendo estimado con certeza lo que permitiría la convergencia del método hacia una política de operación mejorada.

Tabla 5. 10: Intervalos de confianza (95 percentilo) y valor más frecuente de los parámetros de cada modelo de la biblioteca de modelos de tendencias re-estimados en las experiencias 1 y 2

Parámetro	Experiencia 1				Experiencia 2			
	Monod	Contois	Manten,	2° orden	Monod	Contois	Manten,	2° orden
μ_{max}	[0,7005-4,8373] 0,7212		[0,8791-0,9427] 0,8959	[2,9517-3,9646] 3,6388	[1,2241-4,8306] 3,6945	[1,0746-3,4714] 1,1345		
K_s					[1,4553-30,1743] 30,0307		[2,9719-3,3085] 2,9765	
Y_{xs}	[0,3414-1,0000] 0,9967	1,0000		[0,6765-0,9787] 0,8080				
K_I	-	-		-				[44,4242-64,4593] 64,3591
k_{lis}	[0,0074-0,1430] 0,1423	[0,2977-0,3706] 0,3207		[0,0120-0,1390] 0,1117		[0,2977-0,3646] 0,2980		[0,0803-0,1390] 0,1387
m			[0,0922-0,1954] 0,1742					

Tabla 5. 11: Intervalos de confianza (95 percentilo) y valor más frecuente de los parámetros de cada modelo de la biblioteca de modelos de tendencias re-estimados en la experiencia 3

Parámetro	Experiencia 3			
	Monod	Contois	Manten.	2° orden
μ_{max}		[1,0746-3,3959] 1,0862		
K_s		[0,0602-0,8413] 0,0642		
Y_{xs}	1,0000		[0,6369-0,8880] 0,6382	[0,8285-0,9787] 0,9787
K_I				0,0803
k_{lis}				
m			[0,1317-0,1954] 0,1951	

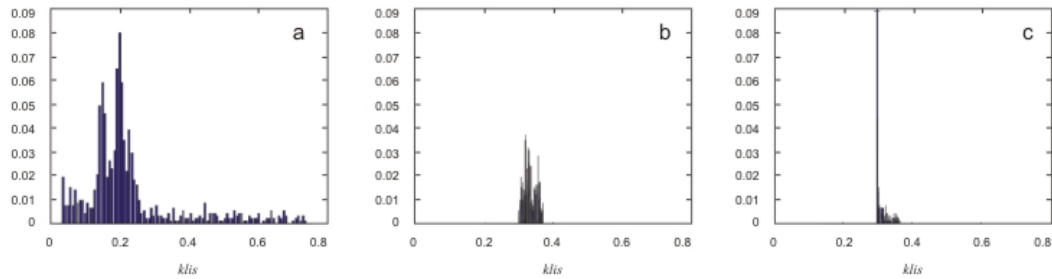


Figura 5.10: *fdp* empíricas para el parámetro k_{lis} del modelo Contois estimadas luego de las experiencias inicial (a), 1 (b) y 2 (c).

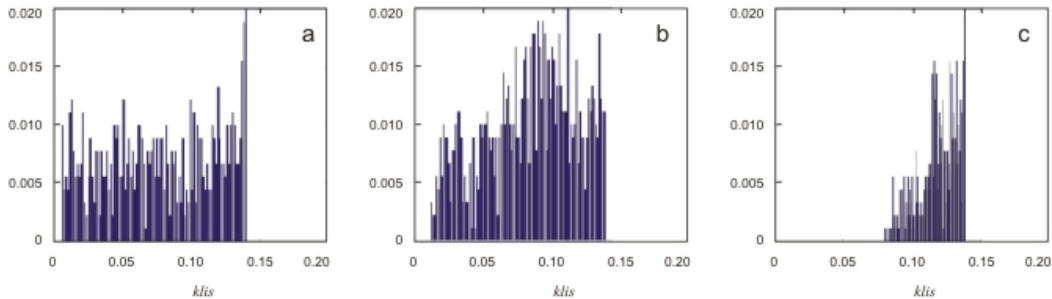


Figura 5.11: *fdp* empíricas para el parámetro k_{lis} del modelo 2° orden estimadas luego de las experiencias inicial (a), 1 (b) y 2 (c).

Para conocer la estabilidad de la convergencia del método propuesto ha sido realizada una experiencia extra en la que fue implementado el perfil de alimentación de la Experiencia 3 pero con la siguiente parametrización modificada: t_{feed} : 9,8 h; $t_{feedend}$: 12,5 h; S_0 : 10,34 g.L⁻¹; S_{feed} : 29,5 g.L⁻¹. El valor de J obtenido fue de 14,7 gDCW. Dado que las condiciones operativas no estaban tan alejadas de las implementadas en la experiencia 3, realizar el *GSA* se obtuvo el mismo perfil de sensibilidad de J . La re-estimación de los parámetros más relevantes de los modelos de tendencias resultó en los mismos en intervalos de confianza ligeramente diferentes pero los valores más frecuentes coincidieron con los estimados con los datos experimentales de la experiencia 3. Debido a lo anterior, el diseño del próximo experimento teniendo en cuenta los datos de la experiencia extra coincide con la experiencia 4 $w_4^{2^\circ orden}$ de la tabla 5.5 con lo que puede decirse que la convergencia del método es estable ante variaciones razonables en los parámetros de la política de operación.

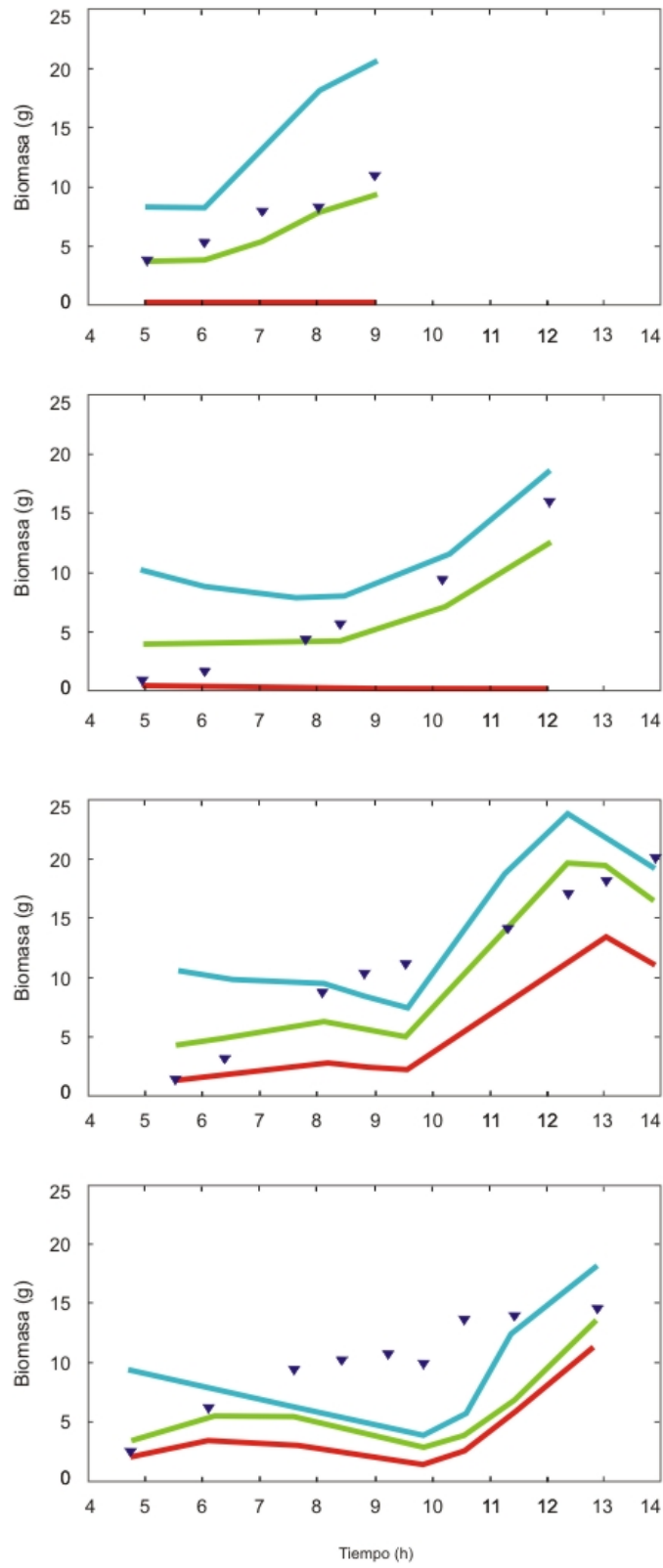


Figura 5.12: Evolución de espacio de estimaciones de J del modelo Contois durante la optimización

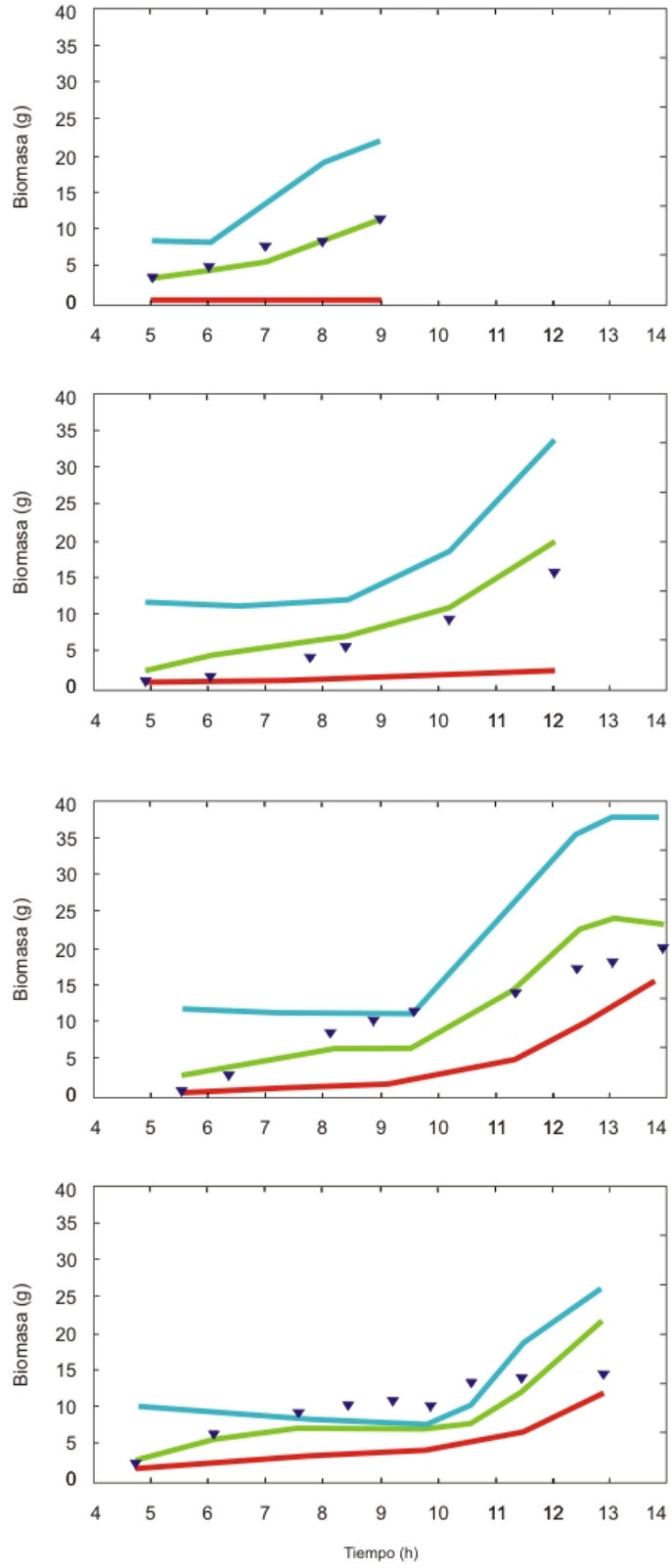


Figura 5.13: Evolución de espacio de estimaciones de J del modelo 2º orden durante la optimización

5.6 Conclusiones:

La metodología para el Diseño de Experimentos Dinámicos basado en Modelos para Optimización ha sido implementada para la mejora de la producción de *E. coli* BL21 transformada con la construcción pET-24b-*bm95* en un biorreactor a escala banco.

La optimización estuvo asistida por una biblioteca con 4 modelos de tendencias que fueron siendo sesgados iterativamente para una mejor estimación de la política mejorada. Además, fueron investigados y definidos los rangos de variación posibles de los parámetros que definen el espacio de políticas de operación implementables W en para el caso de estudio.

De los resultados obtenidos puede concluirse que:

- La metodología para el Diseño de Experimentos Dinámicos basado en Modelos para Optimización ha probado se útil en la mejora experimental de la producción de *E. coli* en un cultivo *fed-batch*.
- La metodología converge a una política de operación del biorreactor mejorada luego del diseño de 3 experiencias para optimización que al ser implementada produce una masa celular de casi el doble comparada con el inicio del proceso de optimización [20,3 gDCW lo que significaría una producción estimada de aproximadamente 522 mg de rBm95 (equivalente a 251 dosis de una formulación vacunal)].
- Durante la mejora iterativa, los espacios de incertidumbre paramétrica de los modelos de tendencias se vieron reducidos gracias a la re-estimación de los parámetros que más repercusión tenían sobre la variabilidad del índice de *performance* según los *GSA* realizados en cada condición operativa. Sin embargo, al final de la optimización los modelos de tendencias están tan restringidos respecto a sus posibles parametrizaciones que no son capaces de predecir correctamente la trayectoria completa de J .
- La estabilidad de la convergencia del método respecto a la política óptima ha sido evaluada mediante la implementación de una variante de la política $w_3^{Contois}$. Luego de rehacer tanto el *GSA* como la re-estimación de los parámetros más relevantes utilizando los datos experimentales recolectados, la política óptima recalculada coincide con la política mejorada $w_4^{2^o\text{orden}}$.

5.7 Bibliografía

- Babu, K.R., Seaminathan, S., Marten, S., Khanna, N., Rinas, U. (2000). Production of interferon in high cell density cultures of recombinant *Escherichia coli* and its single step purification from refolded inclusion body proteins. *Applied Microbiology and Biotechnology* **53**, 655–660.
- Beom S., Seung C., Sang Y., Yong K., Ho N. (2004); High cell density fed-batch cultivation of *Escherichia coli* using exponential feeding combined with pH-stat; *Bioprocess Biosyst Eng.* **26**: 147–150.
- Bhattacharya y Dubey (1997); Effects of dissolved oxygen and oxygen mass transfer on overexpression of target gene in recombinant *E. coli*; *Enzyme and Microbial Technology* **20**. 355-360.
- Chao, Y.P., Law, W.S., Chen, P.T., Hung, W.B. (2002). High production of heterologous proteins in *Escherichia coli* using the thermo-regulated T7 expression system. *Applied Microbiology and Biotechnology* **58**, 446–453.
- Choi, J.H., Jeong, K.J., Kim, S.C., Lee, S.Y. (2000). Efficient secretory production of alkaline phosphatase by high cell density culture of recombinant *Escherichia coli* using the *Bacillus* sp. endoxylanase signal sequence. *Applied Microbiology and Biotechnology* **53**, 640–645.
- Choi, J., Keum K. and Lee S. (2006); Production of recombinant proteins by high cell density culture of *Escherichia coli*. Review; *Chemical Engineering Science* **61**. 876 – 885.
- Datar RV, Cartwright T and Rosen C-G. (1993). Process economics of animal cell and bacterial cell fermentations: A case study analysis of tissue plasminogen activator. *Bio/Technology* **11**, 349–357.
- el-Mansi EM, Holms WH. (1989). Control of carbon flux to acetate excretion during growth of *Escherichia coli* in batch and continuous cultures. *J Gen Microbiol.*; **135**(11):2875–2883.
- Enfors , S.O., Jahic, M., Rozkov, A., Xu, B., Hecker, M., Jurgens, B., Kruger, E., Schweder, T., Hamer, G., O’Beirne, D., and 22 other authors (2001). Physiological responses to mixing in large scale bioreactors. *J. Biotechnol.*, **85**, 175–185.
- Eppstein, L., ShevJtz, J., Yang, X-M. and Weiss, S. (1989); Increased biomass production in a benchtop fermentor ; *Bio/Tedmology*, **7**, 1178 1181.
- Ettler, P. (1990). Scale-up and scale-down techniques for fermentations of polyene antibiotics. *Collect. Czech. Chem. Commun.*, **55**, 1730–1740.
- Fieschko JC. (1989). Fermentation technology using recombinant microorganisms. In: Rehm HJ, Reed G (eds) *Biotechnology*, vol 7b. VCH, Weinheim, 117-140
- Gerigk, M., Bujnicki, R., Ganpo-Nkwenkwa, E., Bongaerts, J., Sprenger, G., Takors, R., (2002). Process control for enhanced L-phenylalanine production using different recombinant *Escherichia coli*. *Biotechnology and Bioengineering* **80**, 746–754.
- Han, K., Lim, H.C, Hong, J. (1992). Acetic acid formation in *Escherichia coli* fermentation. *Biotech. Bioeng.* **39**, 663–671
- Hellmuth K., Korz D.J., Sanders E.A. and Deckwer W.D. (1994); Effect of growth rate on stability and gene expression of recombinant plasmids during continuous and high

- cell density cultivation of *Escherichia coli* TG1; *Journal of Biotechnology*. **32**(3), 289-298.
- Holms WH. (1986). The central metabolic pathways of *Escherichia coli*: relationship between flux and control at a branch point, efficiency of conversion to biomass, and excretion of acetate. *Curr Top Cell Regul.*; **28**.69–105.
- Horn U, Strittmatter W, Krebber A, Knupfer U, Kujau M, Wenderoth R, y col. (1996). High volumetric yields of functional dimeric miniantibodies in *Escherichia coli*, using an optimized expression vector and high-cell density fermentation under non-limited growth conditions. *Appl Microbiol Biotechnol*; **46**.524–32.
- Hu, S.Y., Wu, J.L., Huang, J.H. (2004). Production of tilapia insulin-like growth factor-2 in high cell density cultures of recombinant *Escherichia coli*. *Journal of Biotechnology* **107**, 161–171.
- Jensen B. and Carlsen S. (1990), Production of recombinant human growth hormone in *Escherichia coli*: expression of different precursors and physiological effects of glucose, acetate, and salts, *Biotechnology and Bioengineering* **36**, 1–11
- Jeong, K.J., Lee, S.Y. (1999). High-level production of human leptin by fedbatch cultivation of recombinant *Escherichia coli* and its purification. *Applied and Environmental Microbiology* **65**, 3027–3032.
- Jeong, K.J., Choi, J.H., Yoo, W.M., Keum, K.C., Yoo, N.C., Lee, S.Y., Sung, M.-H., (2004). Constitutive production of Human Leptin by Fed-batch Culture of Recombinant rpoS– *Escherichia coli*. *Protein Expression and Purification* **36**, 150–156.
- Jung G, Denèfle P, Becquart J, Mayaux JF. (1988); High-cell density fermentation studies of recombinant *Escherichia coli* strains expressing human interleukin-1 beta; *Ann Inst Pasteur Microbiol*. **139**(1):129-46.
- Kane, J.F., Hartley, D.L., (1988). Formation of recombinant protein inclusion bodies in *Escherichia coli*. *Trends in Biotechnology* **6**, 95–101.
- Kazan D., Çamurdan A. and Hotaçsu A. (1995); The effect of glucose concentration on the growth rate and some intracellular components of a recombinant *E. coli* culture; *Process Biochemistry*; **30**(3), 269-273.
- Kleist, S., Miksch, G., Hitzmann, B., Arndt, M., Friehs, K., Flaschel, E. (2003). Optimization of the extracellular production of a bacterial phytase with *Escherichia coli* by using different fed-batch fermentation strategies. *Applied Microbiology and Biotechnology* **61**, 456–462.
- Kleman GL, Strohl WR. (1994); Developments in high cell density and high productivity microbial fermentation. *Curr Opin Biotechnol*. **5**(2):180–186.
- Konstantinov K, Nishio N. and Yoshida T. (1990); Glucose feeding strategy accounting for the decreasing oxidative capacity of recombinant *Escherichia coli* in fed-batch cultivation for phenylalanine production; *Journal of Fermentation and Bioengineering*. **70**(4), 253-260.
- Korz D.J., Rinas U., Hellmuth K., Sanders E.A., Deckwer W.D. (1995); Simple fed-batch technique for high cell density cultivation of *Escherichia coli*; *Journal of Biotechnology* **39**, 59-65.
- Lee S. Y. y Chang H. N.(1993); High cell density cultivation of *Escherichia coli* W using sucrose as a carbon source; *Biotechnology letters*; **15**(9); 971-974.

- Lee S. y Chang H. (1994); Effect of complex nitrogen source on the synthesis and accumulation of poly(3-hydroxybutyric acid) by recombinant *Escherichia coli* in flask and fed-batch cultures; *Journal of Environmental Polymer Degradation*; **2**(3), 169-177.
- Lee, S.Y. (1996). High cell density cultivation of *Escherichia coli*. *Trends in Biotechnology* **14**, 98–105.
- Luli y Strohl (1990), Comparison of Growth, Acetate Production, and Acetate Inhibition of *Escherichia coli* Strains in Batch and Fed-Batch Fermentations, *App and Environ Microbiol*, **56**(4) 1004-1011.
- MacDonald H. and Neway J. (1990); Effects of medium quality on the expression of human interleukin-2 at high cell density in fermentor cultures of *Escherichia coli* K-12; *Appl Environ Microbiol*. **56**(3): 640–645.
- Makrides, S.C. (1996). Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiological Review* **60**, 512–538.
- Märkl H., Zenneck C., Dubach A. Ch. and Ogbonna J. (1993); Cultivation of *Escherichia coli* to high cell densities in a dialysis reactor; *App Microbiol and Biotech*, **39**(1), 48-52.
- Meyer H., Leist C. and Fiechter A. (1984). Acetate formation in continuous culture of *Escherichia coli* K12D1 on defined and complex media, *J. Biotech*. **1**, 355–358.
- Mijts, B.N., Schmidt-Dannert, C. (2003). Engineering of secondary metabolite pathway. *Current Opinion in Biotechnology* **14**, 597–602.
- Mizutani, S., Mori, H., Shimizu, S., Sakaguchi, K. and Kobayashi, T. (1986); Effect of amino acid supplement on cell yield and gene product in *Escherichia coli* harboring plasmid, *Biotechnology and Bioengineering*, **28**(2), 204–209.
- Mori, H., Yano, T., Kobayashi, T. and Shimizu, S. (1979); High cell density cultivation of biomass in fed-batch system with DO-stat; *J. Chem.Eng.Jpn* **12**, 313-319
- Pan JG, Rhee JS, Lebeault JM. (1987). Physiological constraints in increasing biomass concentration of *Escherichia coli* B in fed-batch cultures. *Biotechnology Letters*, **9**(2):89-94.
- Panda A., Khan R.H., Appa Rao K.B.C., Totey S.M. (1999); Kinetics of inclusion body production in batch and high cell density fed-batch culture of *Escherichia coli* expressing ovine growth hormona; *Journal of Biotechnology* **75**. 161–172.
- Qiu, J., Swartz, J.R., Georgiou, G. (1998). Expression of active human tissue-type plasminogen activator in *Escherichia coli*. *App and Environ Microbiol* **64**, 4891–4896.
- Riesenber, D. (1991); High-cell-density cultivation of *Escherichia coli*; *Current Opinion in Biotechnology*; **2**, 3, 380-384.
- Roe A., O’Byrne C., McLaggan D. and Booth I. (2002); Inhibition of *Escherichia coli* growth by acetic acid: a problem with methionine biosynthesis and homocysteine toxicity; *Microbiology* **148**, 2215-2222
- Seo, D. J., Chung, B. H., Hwang, Y. B. and Park, Y. H. (1992); Glucose-limited fed-batch culture of *Escherichia coli* for production of recombinant human interleukin-2 with the DO-stat method; *Journal of fermentation bioengineering* **74**(3), 196-198.
- Shiloach J., Fass R. (2005); Growing *E. coli* to high cell density – A historical perspective on method development; *Biotechnology Advances* **23**; 345-357.

- Shimizu M., Iijima S. and Kobayash T. (1992); Production of insecticidal protein of *Bacillus thuringiensis* by cultivation of recombinant *Escherichia coli*; *Journal of Fermentation and Bioengineering*. **74**(3), 163-168.
- Shin, C.S., Hong, M.S., Bae, C.S., Lee, J.W. (1997). Enhanced production of human mini-proinsulin in fed-batch culture at high density of *Escherichia coli* BL21(DE3)[pEET-3aT2M2]. *Biotechnology progress* **13**, 249–257.
- Sletta, H., Nedal, A., Aune, T.E.V., Hellebust, H., Hakvag, S., Aune, R., Ellingsen, T.E., Valla, S., Brautaset, T. (2004). Broad-host-range plasmid pJB658 can be used for industrial-level production of a secreted host toxic single-chain antibody fragment in *Escherichia coli*. *Applied and Environmental Microbiology* **70**, 7033–7039.
- Strandberg L., Andersson L. and Sven-Olof Enfors (1994); The use of fed batch cultivation for achieving high cell densities in the production of a recombinant protein in *Escherichia coli*; *FEMS Microbiology Reviews*; **14**(1), 53-56.
- Sweere, A. P. J., Luyben, K. C. A. M., and Kossen, N.W. F. (1987). Review. Regime analysis and scale-down: tools to investigate the performance of bioreactors. *Enzyme Microb. Technol.*, **9**, 386–398
- Trilli, A. (1986). Scale-up of fermentation, p. 277–307. In Demain, A. L. and Solomon, N. A. (ed.), *Manual of industrial microbiology and biotechnology*. American Society for Microbiology, Washington, D.C..
- Varma A. and Palsson B. (1994); Stoichiometric Flux Balance Models quantitatively Predict Growth and Metabolic By-Product Secretion in Wild-Type *Escherichia coli* W3110; *App and Environ Microbiol* **60**(10), 3724-3731.
- Wong, H.H., Kim, Y.C., Lee, S.Y., Chang, H.N. (1998). Effect of postinducing nutrient feeding strategies on the production of bioadhesive protein in *Escherichia coli*. *Biotechnology and Bioengineering* **60**, 271–276.
- Yee L, Blanch HW (1992). Recombinant protein expression in high cell density fed-batch cultures of *Escherichia coli*. *Biotechnology (N Y)*, **10**:1550-1556.
- Yoon S., Kang W., Park T. (1994); Fed-batch operation of recombinant *Escherichia coli* containing trppromoter with controlled specific growth rate; *Biotechnology and Bioengineering* **43**(10), 995–999.
- Yuan, H., Yang, X., Hua, Z.C. (2004). Optimization of expression of an annexin V-hirudin chimeric protein in *Escherichia coli*. *Microbiological Research* **159**, 147–156.
- Yim, S.C., Jeong, K.J., Chang, H.N., Lee, S.Y. (2001). High level secretory production of human G-CSF by fed-batch culture of recombinant *Escherichia coli*. *Bioprocess and Biosystems Engineering* **24**, 249–254.
- Young, T. B. (1979). Fermentation scale-up: industrial experience with a total environmental approach. *Ann. N.Y. Acad. Sci.* **326**, 165–180.

*...random is very difficult to achieve,
organization always merges back if you don't pay attention...*

*Stephanie a Stephane
en Soñando Despierto*

Capítulo 6

Conclusiones y Perspectivas futuras

6.1 Conclusiones generales

En el presente trabajo de tesis han sido estudiadas, utilizadas e implementadas herramientas conceptuales y metodológicas de campos del conocimiento tales como matemática, estadística y teoría de la probabilidad tanto desde una perspectiva clásica como desde una punto de vista bayesiano, modelado cinético de reacciones complejas, *process system engineering*, biología, bioquímica, microbiología, biología celular y molecular, entre otras; haciendo del mismo un interesante caso de trabajo multidisciplinario.

Como aporte académico reivindico particularmente que una metodología completa para el Diseño de Experimentos Dinámicos basado en Modelos Probabilísticos ha sido formulado e implementado para la mejora del índice de *performance* de casos de estudio biotecnológicos tanto *in silico* como reales.

Dicho algoritmo ha sido estudiado desde el punto de vista teórico mediante el análisis de las condiciones que deben cumplirse para garantizar la convergencia tanto: del método a una política de operación del proceso bajo estudio como de esa política mejorada a la política de operación óptima. Además, en el caso de estudio real en que ha sido mejorada la producción de rBm95 en *Escherichia coli* también ha sido evaluada la estabilidad de la convergencia a variaciones en la implementación de la política de operación.

El método se compone de 3 etapas bien diferenciadas:

1. Diseño de un experimento dinámico basado en una biblioteca de modelos probabilísticos. Dicho experimento se compone de una política de operación del proceso y de una estrategia de muestreo optimizada.
2. Implementación del experimento dinámico para optimización.
3. Mejora de las herramientas utilizadas para la estimación del óptimo, es decir, la biblioteca de modelos probabilísticos. Esto es realizado procurando hacer una utilización óptima de la información contenida en los datos experimentales.

Durante la formulación de la metodología para la optimización experimental iterativa han sido implementadas herramientas numéricas como el *GSA* y *bootstrapping*. Lo cual resultó en la propuesta de un algoritmo de discriminación de modelos de primeros principios y en la posibilidad de agregar luz sobre las posibles etapas de reacción que ocurren durante una reacción química regioselectiva. Además, con los resultados

obtenidos ha podido ser modelada la etapa de síntesis no isotérmica de ivermectinas de la planta de producción de la empresa OVER SRL.

Como aportes tecnológicos de este trabajo puede reivindicarse la obtención de un clon de *E. coli* transformado genéticamente con la construcción pET-24+-*bm95*. Dicho microorganismo recombinante es capaz de expresar satisfactoriamente la proteína rBm95 utilizada para la formulación de vacunas contra la garrapata *Rhipicepalus microplus*.

Además, el Diseño de Experimentos Dinámicos para Optimización ha sido implementado en el cultivo de dicho microorganismo transformado genéticamente. Luego del diseño de 3 experimentos, el método converge a una política de operación mejorada del biorreactor a escala banco que permite una producción del doble de biomasa del obtenido al inicio de las experiencias que equivale a 251 dosis vacunales.

6.2 Perspectivas futuras

Como proyecciones académicas podría señalarse el estudio del comportamiento de la metodología del Diseño de Experimentos Dinámicos para Optimización basado en Modelos de Tendencias Probabilísticos respecto a:

- la dependencia de la convergencia y dirección de búsqueda de la política de operación óptima de la experiencia inicial ya que con los datos experimentales obtenidos en ella son definidos los espacios de incertidumbre paramétricas de los modelos de tendencias de la biblioteca de modelos para optimización y en suma las posibilidades de estimación del índice de *performace*.
- la incorporación de una perspectiva bayesiana en la que es tomada en cuenta la presencia de perturbaciones experimentales que pueden afectar tanto la implementación como la adquisición de los datos experimentales.

El trabajo futuro desde el punto de vista tecnológico debería centrarse en la mejora tanto de las densidades de biomasa obtenidos en cultivos de *E. coli* como de rBm95. Además, podría cambiarse el vector de expresión por *Picchia pastoris* que ha demostrado tener altos rendimientos en la producción de la proteína de interés.

Durante el presente trabajo también han sido determinadas experimentalmente las evoluciones del metabolito acetato en el medio de cultivo. Esta variable resultó ser altamente dinámica como para incorporarla en un esquema como el presentado donde los modelos de tendencias de la biblioteca de modelos no contaban con más de 5 parámetros.

Por lo anterior, la incorporación de la estimación de la variable acetato en las herramientas de optimización podría redundar en la mejora de la producción de rBm95 habida cuenta de los efectos metabólicos que causa en *E. coli*.

Además, debido al alto costo del inductor utilizado, el *scaling up* de la producción de rBm95 debe contemplar la posibilidad de utilizar otra cepa para la expresión de dicha proteína. Hasta el momento han sido realizados ensayos preliminares con la cepa *E. coli* MP101 cuya principal característica es que presenta altos niveles de expresión cuando es utilizada lactosa como señal metabólica de inducción.

ANEXOS

Anexo I: ANALISIS DE SENSIBILIDAD GLOBAL (GSA)

Teoría

Sea la representación matemática o computacional del modelo que desea ser estudiado

$$Y = f(\theta_i), i = 1, 2, \dots, k \quad (I.1)$$

donde θ_i s son los parámetros del modelo, los cuales se comportan como variables aleatorias en su rango de variación $[\theta_i^{\min}, \theta_i^{\max}]$, no están correlacionados y las funciones de densidad de probabilidad $p_i(\theta_i)$ individuales son conocidas.

Entonces, la función de densidad de probabilidad conjunta puede ser representada como:

$$P(\theta_i) = \prod_{i=1}^k p_i(\theta_i) \quad (I.2)$$

Esa información define un espacio de incertidumbre Θ que repercutirá sobre la variabilidad en las estimaciones o respuesta Y del modelo. Entonces la esperanza $E(Y)$ y varianza $V(Y)$ pueden ser calculados como:

$$E(Y) = \iint \dots \int f(\theta_1, \dots, \theta_k) \prod_{i=1}^k p_i(\theta_i) d\theta_i \quad (I.3)$$

$$\begin{aligned} V(Y) &= \iint \dots \int (f(\theta_1, \dots, \theta_k) - E(Y))^2 \prod_{i=1}^k p_i(\theta_i) d\theta_i \\ &= \iint \dots \int f^2(\theta_1, \dots, \theta_k) \prod_{i=1}^k p_i(\theta_i) d\theta_i - E^2(Y) \end{aligned} \quad (I.4)$$

Si uno de los factores, digamos θ_j , es fijado en un valor $\tilde{\theta}_j \in [\theta_j^{\min}, \theta_j^{\max}]$, entonces la varianza condicionada de Y dado que θ_j es conocido será calculada como:

$$V(Y|\theta_j = \tilde{\theta}_j) = \iint \dots \int f^2(\theta_1, \dots, \tilde{\theta}_j, \dots, \theta_k) \prod_{\substack{i=1 \\ i \neq j}}^k p_i(\theta_i) d\theta_i - E^2(Y|\theta_j = \tilde{\theta}_j) \quad (I.5)$$

y esa varianza condicionada es una buena medida de la sensibilidad de Y respecto de θ_j . Sin embargo, dicha medida depende del valor puntual $\tilde{\theta}_j$ en el que θ_j ha sido fijado para el cálculo. Además existen casos de modelos no lineales en los que $V(Y|\theta_j = \tilde{\theta}_j) \geq V(Y)$. Por lo tanto, para eliminar esa dependencia se realiza la

integración de $V(Y|\theta_j = \tilde{\theta}_j)$ sobre $p_j(\theta_j)$, la función de densidad de probabilidad (*fdp*) de θ_j :

$$E(V(Y|\theta_j)) = \iint \dots \int f^2(\theta_1, \dots, \theta_j, \dots, \theta_k) \prod_{i=1}^k p_i(X_i) dX_i \dots \dots - \int E^2(Y|\theta_j = \tilde{\theta}_j) p_j(\tilde{\theta}_j) d\tilde{\theta}_j \quad (I.6)$$

Si a la varianza total se le sustrae este último estadístico se obtiene:

$$V(Y) - E(V(Y|\theta_j)) = \int E^2(Y|\theta_j = \tilde{\theta}_j) p_j(\tilde{\theta}_j) d\tilde{\theta}_j - E^2(Y) \quad (I.7)$$

El lado izquierdo de la Ec. (I.7) es la varianza de Y dado que θ_j es conocido:

$$V(Y|\theta_j) = V(Y) - E(V(Y|\theta_j)) \quad (I.8)$$

y puede ser utilizado como una buena medida de la sensibilidad de Y respecto de θ_j .

$V(Y|\theta_j)$ es llamado como el efecto de primer orden de θ_j sobre Y . Como corolario se tiene que $V(Y|\theta_j) \leq V(Y)$.

Si el efecto de primer orden obtenido es dividido por la varianza incondicionada (total) de Y es obtenido el índice de sensibilidad global de efecto principal:

$$S_j = \frac{V(Y|\theta_j)}{V(Y)} \quad (I.9)$$

el cual, por el corolario está definido en el intervalo $[0,1]$.

Otro estadístico importante es el definido como $E(V(Y|\theta_{-j}))$ que representa la varianza de Y que sería explicada si se conocieran exactamente todos los factores excepto θ_j , esto es:

$$E(V(Y|\theta_{-j})) = \sum_{m=1}^k V_{jm} = V(Y) - V(Y|\theta_j) \quad (I.10)$$

lo cual implica la suma de todos los términos de la descomposición de la varianza de Y en los que aparece θ_j . Esto lleva a la definición de otro índice de sensibilidad llamado índice de efectos totales S_{T_j} :

$$S_{T_i} = \frac{E(V(Y|\theta_{-j}))}{V(Y)} = 1 - \frac{V(E(Y|\theta_{-j}))}{V(Y)} = S_i + \sum_j S_{ij} + \sum_j \sum_k S_{ijk} + \dots \quad (\text{I.11})$$

Sobol' (Sobol', 2001) ha propuesto un esquema de descomposición de la varianza $V(Y)$ de la forma:

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \dots + V_{12\dots k} \quad (\text{I.12})$$

donde:

$$V_i = V(E(Y|\theta_i)) \quad (\text{I.13})$$

$$V_{ij} = V(E(Y|\theta_i, \theta_j)) - V_i - V_j \quad (\text{I.14})$$

Así, la $V(Y)$ puede descomponerse en k términos de primer orden V_i ; $\frac{k(k-1)}{2}$ términos de segundo orden V_{ij} que representa a la varianza de Y explicada por la interacción de θ_i y θ_j ; y así sucesivamente, hasta el último término de orden k haciendo un total de $2^k - 1$ términos. Sobol' además ha probado la unicidad de la descomposición si los factores son independientes y como corolario se tiene:

$$1 = \sum_i S_i + \sum_i \sum_{j>i} S_{ij} + \dots + S_{12\dots k} \quad (\text{I.15})$$

Algunas de las consecuencias de los resultados anteriores son:

- Si $\sum_i S_i = 1$ entonces el modelo estudiado es completamente aditivo y $Y = f(\theta_i), i = 1, 2, \dots, k$ puede descomponerse como la suma de k funciones cada una siendo únicamente función de un factor. Lo que nos da un criterio para definir un modelo no aditivo como $\sum_i S_i \leq 1$.
- Si $S_{T_i} \equiv 0$ entonces tenemos la condición necesaria y suficiente para asegurar que el factor θ_i no influye en la varianza de Y ya que eso implica que $E(V(Y|\theta_{-j})) \equiv 0$, lo cual sólo es posible si $V(Y|\theta_{-j}) = 0 \forall \theta_i \in [\theta_i^{\min}, \theta_i^{\max}]$ ya que por definición $V(\bullet) \geq 0$.
- La diferencia $S_{T_i} - S_i$ da una idea del grado de interacción del factor θ_i con el conjunto θ_{-i} .

- Los índices S_i pueden ser utilizados para construir un *ranking* u ordenamiento jerárquico en el que los factores de variabilidad de un modelo sean ordenados según el efecto que producen en la varianza de la salida debido a que no se conocen exactamente. Esta técnica es conocida como priorización de parámetros.
- Los resultados obtenidos al dividir el conjunto de factores en θ_i y θ_{-i} puede ser extendido para una división que contenga más de un factor en cada caso:

$$\boldsymbol{\theta} = \boldsymbol{\theta}_u \cup \boldsymbol{\theta}_v; \mathbf{u} = 1, 2, \dots, r, \mathbf{v} = 1, 2, \dots, k - r.$$

La varianza $V(Y)$ debida al subconjunto $\boldsymbol{\theta}_u$ puede ser definida como:

$$V(\boldsymbol{\theta}_u) = \sum_{u=u_1}^{u_m} V_u + \sum_{u=u_1}^{u_m} \sum_{\substack{f=f_1 \\ f \neq u}}^{f_m} V_{uf} + \dots + V_{u_1, u_2, \dots, u_m} \quad (\text{I.16})$$

Lo mismo puede ser realizado para el subconjunto $\boldsymbol{\theta}_v$, de forma que la varianza total de Y debida al subconjunto $\boldsymbol{\theta}_u$ es:

$$V_{\boldsymbol{\theta}_u}^{tot} = V - V_{\boldsymbol{\theta}_v} \quad (\text{I.17})$$

De esta forma, dos índices de sensibilidad global adicionales pueden ser definidos:

$$S_{\boldsymbol{\theta}_u} = \frac{V_{\boldsymbol{\theta}_u}}{V} \quad (\text{I.18})$$

$$S_{\boldsymbol{\theta}_u}^{tot} = \frac{V_{\boldsymbol{\theta}_u}^{tot}}{V} = 1 - \frac{V_{\boldsymbol{\theta}_v}}{V} \quad (\text{I.19})$$

Si $S_{\boldsymbol{\theta}_u}$ es lo suficientemente pequeño como para considerar que el aporte a la varianza total del modelo de los factores $\theta_i \in \boldsymbol{\theta}_u$ es despreciable, entonces $\boldsymbol{\theta}_u$ puede fijarse en cualquier valor de su dominio. Este resultado es interesante para reducir la dimensionalidad del modelo en estudio. En la próxima sección será mostrado que el error cometido al utilizar un “modelo reducido” por la fijación de algunos de sus parámetros permanece acotado y su cota máxima puede ser calculada.

Error esperado debido a la fijación de parámetros en un modelo

Sea $h(\theta)$ una función de cuadrado integrable utilizada como una aproximación de $f(\theta)$. Una estimación del error de aproximación puede ser obtenida al calcular la distancia escalada:

$$\delta(f, h) = \frac{1}{V} \int [f(\theta) - h(\theta)]^2 d\theta \quad (I.20)$$

Supongamos que $S_{\theta_v}^{tot} \ll 1$, por lo tanto Y depende principalmente de θ_u . Si $h(\theta_u) = f(\theta_u, \theta_v^*)$, donde θ_v^* es el vector θ_v cuyos componentes han sido fijados en sus valores nominales, es propuesta como aproximación de $f(\theta)$. Entonces el siguiente teorema nos dice que el error de aproximación $\delta(f, h) = \delta(\theta_v^*)$ depende de $S_{\theta_v}^{tot}$.

Teorema 1 (Sobol', 2007)

Para un $\delta(\theta_v^*)$, $\theta_v^* \in \Theta_v$ arbitrario

$$\delta(\theta_v^*) \geq S_{\theta_v}^{tot} \quad (I.21)$$

Pero si θ_v^* es un punto aleatorio uniformemente distribuido en Θ_v , entonces para un $\varepsilon > 0$:

$$P\left\{\delta(\theta_v^*) < \left(\frac{1}{1+\varepsilon}\right) S_{\theta_v}^{tot}\right\} \geq 1 - \varepsilon \quad (I.22)$$

Otra posible aproximación de $f(\theta)$ es aquella en la que, tras asumir la representación de Sobol', los términos de mayor orden son eliminados ya que como ha sugerido Rabitz (1998) en la mayoría de los modelos el mayor impacto sobre la respuesta es producida por los términos de bajo orden. Entonces:

$$h_L(\theta) = f_0 + \sum_{s=1}^L \sum_{i_1 < \dots < i_s}^n f_{i_1 \dots i_s}(\theta_{i_1}, \dots, \theta_{i_s}) \quad (I.23)$$

con $L \ll n$.

Teorema 2 (Sobol', 2007)

Si el modelo $f(\theta)$ es aproximado por $h_L(\theta)$, entonces el error de aproximación es:

$$\delta(f, h_L) = 1 - \sum_{s=1}^L \sum_{i_1 < \dots < i_s}^n S_{i_1 \dots i_s} \quad (I.24)$$

Prueba:

$$f(\theta) - h_L(\theta) = \sum_{s=L+1}^n \sum_{i_1 < \dots < i_s}^n f_{i_1 \dots i_s}(\theta_{i_1}, \dots, \theta_{i_s}) \quad (I.25)$$

Integrando el cuadrado de esa diferencia e integrando sobre Θ y dividiendo por V se obtiene:

$$\delta(f, h_L) = \sum_{s=L+1}^n \sum_{i_1 < \dots < i_s} S_{i_1 \dots i_s} = 1 - \sum_{s=1}^L \sum_{i_1 < \dots < i_s} S_{i_1 \dots i_s} \quad (I.26)$$

Análisis de Sensibilidad para modelos con parámetros correlacionados

El tratamiento formal del *GSA* y más precisamente la formulación de los métodos de cálculo de estimadores confiables de los índices de sensibilidad para los casos en los que los modelos presentan parámetros correlacionados es incipiente. Si bien en la bibliografía existen trabajos incipientes, en estos los autores tratan ejemplos sencillos y reconocen que sus propuestas no útiles para el uso generalizado sobre todo si son considerados modelos altamente no lineales, cuyos parámetros están altamente correlacionados (Jaques, 2006; Xu y Gertner, 2007; 2008).

Debido a lo anterior, en esta tesis los índices de sensibilidad global han sido calculados siguiendo el método computacional propuesto por Saltelli (Saltelli, 2006)

Algoritmo computacional para el cálculo de los Índices de Sensibilidad Global

A la hora de calcular los índices de sensibilidad nos encontramos con que la formulación (I.7) es poco práctica para su implementación computacional ya que en un esquema Monte Carlo requiere un doble loop:

- uno interno en el que $E^2(Y|\theta_j = \tilde{\theta}_j)$ sea calculada
- otro externo en el que sea realizada la integral sobre $d\tilde{\theta}_j$.

Es por eso que Ishigami y Homma (Ishigami y Homma, 1990) han reescrito la Ec. (I.7) como:

$$\begin{aligned} \int E^2(Y|\theta_j = \tilde{\theta}_j) p_j(\tilde{\theta}_j) d\tilde{\theta}_j &= \\ &= \int \left\{ \int \dots \int f(\theta_1, \dots, \tilde{\theta}_j, \dots, \theta_k) \prod_{\substack{i=1 \\ i \neq j}}^k p_i(\theta_i) d\theta_i \right\}^2 p_j(\tilde{\theta}_j) d\tilde{\theta}_j \\ &= \int \int \dots \int f(\theta_1, \dots, \tilde{\theta}_j, \dots, \theta_k) f(\theta'_1, \dots, \tilde{\theta}_j, \dots, \theta'_k) \prod_{\substack{i=1 \\ i \neq j}}^k (p_i(\theta_i) d\theta_i) \prod_{\substack{i=1 \\ i \neq j}}^k (p_i(\theta'_i) d\theta'_i) p_j(\tilde{\theta}_j) d\tilde{\theta}_j \end{aligned}$$

$$= \int \int \int \dots \int f(\theta_1, \dots, \theta_j, \dots, \theta_k) f(\theta'_1, \dots, \theta_j, \dots, \theta'_k) \prod_{i=1}^k (p_i(\theta_i) d\theta_i) \prod_{\substack{i=1 \\ i \neq j}}^k (p_i(\theta'_i) d\theta'_i) \quad (I.27)$$

La ecuación (I.16) es la $E(F)$ de un conjunto de $(2k-1)$ factores:

$$F(\theta_1, \dots, \theta_j, \dots, \theta_k, \theta'_1, \dots, \theta'_{j-1}, \theta'_{j+1}, \theta'_k) = f(\theta_1, \dots, \theta_j, \dots, \theta_k) f(\theta'_1, \dots, \theta'_{j-1}, \theta'_j, \theta'_{j+1}, \dots, \theta'_k) \quad (I.28)$$

Saltelli (Saltelli, 2006) han propuesto el siguiente método para el cálculo de los índices de sensibilidad global de efectos principales S_i y efectos totales S_{T_i} , y es el método que ha sido adoptado en esta tesis y comprende las siguientes etapas:

- 1- Dos matrices $\mathbf{M}_1, \mathbf{M}_2$ de N muestreos aleatorios de las distribuciones de densidad de probabilidad de los k factores son generadas:

$$\mathbf{M}_1 = \begin{pmatrix} \theta_{11} & \theta_{12} & \dots & \theta_{1k} \\ \theta_{21} & \theta_{22} & \dots & \theta_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{N1} & \theta_{N2} & \dots & \theta_{Nk} \end{pmatrix} \quad \mathbf{M}_2 = \begin{pmatrix} \theta'_{11} & \theta'_{12} & \dots & \theta'_{1k} \\ \theta'_{21} & \theta'_{22} & \dots & \theta'_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \theta'_{N1} & \theta'_{N2} & \dots & \theta'_{Nk} \end{pmatrix}$$

- 2- A partir de \mathbf{M}_1 y \mathbf{M}_2 son definidas $2k$ matrices adicionales $\mathbf{Q}_i, \mathbf{Q}_{T_i}, i=1, 2, \dots, k$:

$$\mathbf{Q}_i = \begin{pmatrix} \theta'_{11} & \dots & \theta_{1j} & \dots & \theta'_{1k} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \theta'_{N1} & \dots & \theta_{Nj} & \dots & \theta'_{Nk} \end{pmatrix} \quad \mathbf{Q}_{T_i} = \begin{pmatrix} \theta_{11} & \dots & \theta'_{1j} & \dots & \theta_{1k} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \theta_{N1} & \dots & \theta'_{Nj} & \dots & \theta_{Nk} \end{pmatrix}$$

- 3- Entonces Y es evaluada utilizando para los factores los valores de cada renglón de las matrices $\mathbf{M}_1, \mathbf{Q}_i, \mathbf{Q}_{T_i}, i=1, 2, \dots, k$ de manera de obtener los vectores de N elementos: $\mathbf{Y}_1 = f(\mathbf{M}_1), \mathbf{Y}_i = f(\mathbf{Q}_i), \mathbf{Y}_{T_i} = f(\mathbf{Q}_{T_i}), i=1, 2, \dots, k$.

- 4- Los índices de sensibilidad global de efectos principales y totales son calculados como:

$$U_i = \frac{1}{N-1} (\mathbf{Y}_1^T \square \mathbf{Y}_i) \quad (I.28)$$

$$U_{T_i} = \frac{1}{N-1} (\mathbf{Y}_1^T \square \mathbf{Y}_{T_i}) \quad (I.29)$$

$$f_0 = \frac{1}{N} \sum_{j=1}^N \mathbf{Y}(\theta_j) \quad (A.1.30)$$

$$V(Y) \rightarrow \frac{1}{N-1} \sum_{j=1}^N \mathbf{Y}^2(\theta_j) - f_0^2 \quad (\text{I.31})$$

$$V_i = V(E(Y|\theta_i)) \rightarrow U_i - f_0^2 \quad (\text{I.32})$$

$$V_{-i} = V(E(Y|\theta_{-j})) \rightarrow U_T - f_0^2 \quad (\text{I.33})$$

$$S_i = \frac{V_i}{V} \quad (\text{I.34})$$

$$S_{T_i} = 1 - \frac{V_{-i}}{V} \quad (\text{I.35})$$

Bibliografia

Borgonovo, E.; Apostolakis, G. E.; Tarantola, S.; Saltelli, A. (2003). Comparison of global sensitivity analysis techniques and importance measures in PSA. *Reliability Engineering and System Safety*, **79**(2), 175-185.

Campolongo, F.; Cariboni, J.; Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling and Software*; **22**; 1509-1518.

Gatelli, D.; S. Kucherenko, Ratto, M., Pagano, A., Tarantola, S. (2009). Calculating first-order sensitivity measures: A benchmark of some recent methodologies. *Reliability Engineering and System Safety*, **94**(7), 1212-1219.

Helton, J.; Davis, F. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering and System Safety*, **81**; 23-69.

Ishigami, T. Homma, T. (1990). An importance quantification technique in uncertainty analysis for computer models. Proceedings of the ISUMA '90, First International Symposium on Uncertainty Modelling and Analysis, University of Maryland, Dic. 3-6.

Iooss, B. and M. Ribatet (2009). "Global sensitivity analysis of computer models with functional inputs." *Reliability Engineering & System Safety* **94**(7): 1194-1204.

Jacques, J.; Lavergne, C.; Devictor, N. (2006). "Sensitivity analysis in presence of model uncertainty and correlated inputs. *Reliability Engineering and System Safety*, **91**(10-11), 1126-1134.

Kucherenko, S.; Rodriguez-Fernandez, M.; Pantelides, C.; Shah, N. (2009); Monte Carlo evaluation of derivative-based global sensitivity measures. *Reliability Engineering and System Safety*; **94**, 1135-1148.

Plischke, E. (2010). An effective algorithm for computing global sensitivity indices. *Reliability Engineering and System Safety*, **95**(4), 354-360.

Saltelli, A., P. Annoni, Azzina I., Campolongo F., Ratto M. and Tarantola S. (2010). "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index." *Computer Physics Communications* **181**(2): 259-270.

- Saltelli, A.; Ratto, M.; Tarantola, S.; Campolongo, F. (2006). Sensitivity analysis practices: Strategies for model-based inference. *Reliability Engineering and System Safety*; **91**, 1109–1125.
- Saltelli, M.R.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Tarantola, S. (2008). *Global Sensitivity Analysis. The Primer*. Southern Gate, Chichester, West Sussex PO19 8SQ, England, John Wiley & Sons Ltd, The Atrium.
- Sobol' I. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*; **55**, 271–280.
- Sobol', I.; Kucherenko, S. (2009). Derivative based global sensitivity measures and their link with global sensitivity indices. *Mathematics and Computers in Simulation*, **79**; 3009–3017.
- Sobol, I. M.; Tarantola, S.; Gatelli, D.; Kucherenko, S.; Mauntz, W. (2007). Estimating the approximation error when fixing unessential factors in global sensitivity analysis. *Reliability Engineering and System Safety*, **92**(7), 957-960.
- Yu, W.; Harris, T. (2009). Parameter uncertainty effects on variance-based sensitivity analysis. *Reliability Engineering and System Safety*, **94**, 596–603.
- Xu, C.; Gertner, G. (2007). Extending a global sensitivity analysis technique to models with correlated parameters. *Computational Statistics and Data Analysis*, **51**, 5579 – 5590.
- Xu, C.; Gertner, G. (2008b). Uncertainty and sensitivity analysis for models with correlated parameters. *Reliability Engineering and System Safety*. **93**; 2008b,1563–1573.
- Ziehn, T. and A. S. Tomlin (2009). "GUI-HDMM - A software tool for global sensitivity analysis of complex models." *Environmental Modelling & Software* **24**(7): 775-785.

Anexo II: *BOOTSTRAPPING*

Teoría

El método de *bootstrap* es utilizado en estadística aplicada para la estimación de parámetros y sus intervalos de confianza. El bootstrap es una técnica computacional, inventada por Efron (1979), que intenta suplir el análisis teórico del enfoque estadístico tradicional con información obtenida a través simulaciones computacionales.

Cuando se necesita conocer un el valor de un parámetro θ , entonces surgen las preguntas:

- ¿Qué estimador $\hat{\theta}$ debería ser utilizado?
- En caso de elegir un $\hat{\theta}$, ¿qué tan exacto es $\hat{\theta}$ como estimador de θ ?

Para responder a la primera pregunta el procedimiento clásico sería obtener una muestra de datos experimentales $\mathbf{y} = (x_i); i = 1, \dots, n$ de observaciones independientes e idénticamente distribuidas (*iid*) del sistema estudiado, es decir que $X_1, X_2, \dots, X_n \square F$ donde F representa la función de densidad de probabilidad (*fdp*), desconocida, del espacio definido por X . Luego θ sería estimado a partir del cálculo de $\hat{\theta}$. La segunda pregunta no sería respondida sino más bien asumida *a priori* mediante la suposición que θ sigue una distribución conocida (normal, *t*-Student, u otra) y los intervalos de confianza para $\hat{\theta}$ serían calculados en función de dicha suposición con una probabilidad $(1 - \alpha) \square 100\%$ de contener al valor verdadero θ .

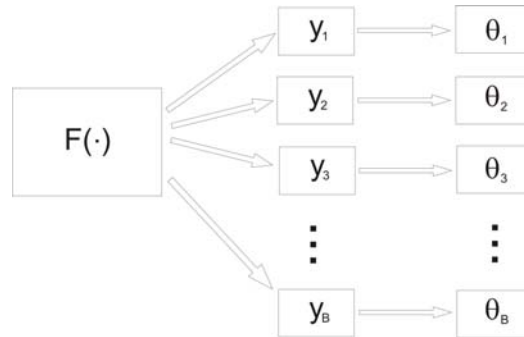
Para responder a la segunda pregunta lo que se necesita es conocer la *fdp* de θ o lo que es lo mismo el error estándar de $\hat{\theta}(\mathbf{y})$.

Sea, el error estándar de $\hat{\theta}$, una función desconocida de la distribución muestreada F que además es función del tamaño de la muestra n y de la forma del estadístico $\hat{\theta}(\mathbf{y})$

$$\sigma(F) = \left[\text{Var}_F \{ \hat{\theta}(\mathbf{y}) \} \right]^{\frac{1}{2}} \quad (\text{II.1})$$

Una estrategia para conocer $\sigma(F)$ podría ser la mostrada en el esquema adjunto en el cual son obtenidas B muestras del sistema desde las que es calculado $\hat{\theta}(\mathbf{y})$ y a partir de

esas estimaciones obtener $\sigma(F)$ o la *fdp* empírica de $\hat{\theta}$. Además, cuando $B \rightarrow \infty$, $\sigma(\hat{\theta}(\mathbf{y}), n) \rightarrow \sigma(\theta)$



Sin embargo, el obtener las muestras *iid* y_1, y_2, \dots, y_B puede resultar extremadamente costoso. Es en esos casos en los que el *bootstrapping* se presenta como una técnica sumamente útil. La idea subyacente detrás del método es que si la única información del sistema bajo estudio es la contenida en una muestra, entonces lo más sensato es intentar conocer el sistema a través de la información contenida en esa muestra más que realizar suposiciones sobre su comportamiento.

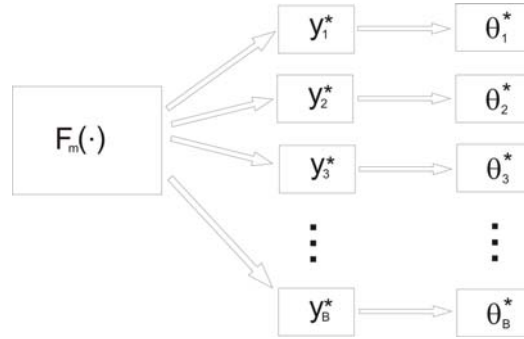
Siguiendo con el ejemplo, con *bootstrapping* es posible calcular mediante simulaciones Monte Carlo estimaciones del error estándar $\hat{\sigma} = \sigma(F_m)$ donde F_m es la distribución empírica de F (considerada su mejor estimador en ausencia de mayor conocimiento).

Sea $\mathbf{y}^* = (x_i^*); i = 1, \dots, n$ una muestra de *bootstrap* tomada de la distribución empírica F_m , esto es \mathbf{y}^* es una muestra de tamaño n tomada con reemplazo desde la muestra real $\{x_1, x_2, \dots, x_n\}$.

El algoritmo Monte Carlo tiene los siguientes tres pasos:

- Generar B muestras de bootstrap aleatorias $\{\mathbf{y}_1^*, \mathbf{y}_2^*, \dots, \mathbf{y}_B^*\}$.
- Para cada \mathbf{y}_i^* calcular el estadístico de interés $\hat{\theta}_i^* = \hat{\theta}(\mathbf{y}_i^*)$.
- Calcular la $\hat{\sigma}_B$ de $\hat{\theta}^*$.

Una representación del método es mostrado en la siguiente ilustración:



Expresado informalmente, para que el método funcione se necesita:

- Que F_m converja a F : esto está garantizado por el teorema de Glivenko-Cantelli (Chernick, 1999) que asegura la convergencia uniforme con probabilidad 1.
- Si $\hat{\theta}$ se refiere a alguna propiedad característica de la distribución (técnicamente esta propiedad es un funcional de las distribuciones de la población), entonces esta característica debe satisfacer ciertas propiedades de suavidad (De Martini, 2008; Barrio, 2002).

Teorema de Glivenko-Cantelli (Glivenko, 1933; Cantelli, 1933).

Asumiendo que X_1, X_2, \dots son variables aleatorias independientes e idénticamente distribuidas (*iid*) en \mathfrak{R} con una *fdp* acumulada $F(x)$. La función distribución empírica para X_1, X_2, \dots, X_n se define como:

$$F_m(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(X_i) \quad (\text{II.2})$$

donde $I_{(-\infty, x]}$ en la función indicadora tal que:

$$I_{(-\infty, x]} = \begin{cases} 1 & X_i \in (-\infty, x] \\ 0 & X_i \notin (-\infty, x] \end{cases} \quad (\text{II.3})$$

Para cada x , $F_m(x)$ es una secuencia de variables aleatorias que converge a $F(x)$ casi seguramente (esto es con probabilidad igual a 1) por la ley de los grandes números, es decir que $F_m(x)$ converge a $F(x)$ puntual y uniformemente:

$$\|F_m(x) - F(x)\|_{\infty} = \sup |F_m(x) - F(x)| \rightarrow 0 \quad (\text{II.4})$$

¿Por qué cada y_i^* tiene la misma cantidad de elementos que el conjunto de datos?

Como fue antes mencionado $\sigma(F) = \sigma(F, n, \hat{\theta})$, es decir que es el error estándar del estadístico $\hat{\theta}$ basado en una muestra de tamaño n tomada de una distribución desconocida F . El estadístico $\hat{\sigma}$ estimado con *bootstrap* es en realidad $\sigma(F, n, \hat{\theta})$ evaluado en $F = \hat{F}$ y el algoritmo de Monte Carlo no converge a $\hat{\sigma}$ si las muestras de bootstrap son diferentes a n .

Intervalos de confianza

En estadística aplicada, el *método estándar* para calcular los intervalos de confianza está basado en asumir la normalidad de la variable aleatoria considerada o garantizando la normalidad mediante la aplicación de alguna transformación (Banks, 2010). Es decir que se espera que el intervalo en el que el valor verdadero de un parámetro será encontrado con una probabilidad de $(1-\alpha) \times 100\%$ está dado por $\langle \hat{\theta} \rangle \pm z_{\alpha/2} \hat{\sigma}_{\theta}$ donde $z_{\alpha/2}$

es el punto de la *pdf* normal que corresponde al $\left(1 - \frac{\alpha}{2}\right) \times 100$ percentilo.

El *bootstrap* también puede otorgarnos los intervalos de variación de un dado parámetro para cierto grado de probabilidad, independientemente de la distribución que dicho parámetro tenga utilizando el *método del percentilo* (DiCiccio, 1996; Efron, 1986).

Sea $\hat{G}(s) = \text{Prob}_* \{ \hat{\theta}^* < s \}$ la *cdf* obtenida de la aplicación del *bootstrap* para calcular

$\hat{\theta}^*$. Entonces el intervalo de confianza buscado queda definido como $\theta \in \left[\hat{G}^{-1}\left(\frac{\alpha}{2}\right), \hat{G}^{-1}\left(1 - \frac{\alpha}{2}\right) \right]$, es decir que el intervalo del método del percentilo es el

intervalo entre los percentilos $\left(\frac{\alpha}{2}\right) \times 100$ y $\left(1 - \frac{\alpha}{2}\right) \times 100$ de la distribución de bootstrap

obtenida para $\hat{\theta}^*$, esto enfatiza el hecho que con esta formulación se intenta dejar una probabilidad de $\left(\frac{\alpha}{2}\right) \times 100$ a cada lado de los extremos y no sólo una $(1-\alpha) \times 100\%$ en

total.

Como es obvio, ambos métodos coinciden si $\hat{\theta} \square N(\theta, \sigma_\theta^2)$, pero el método del percentilo es mejor en aquellos casos en los la distribución de $\hat{\theta}$ es asimétrica ya que logra capturar esa característica.

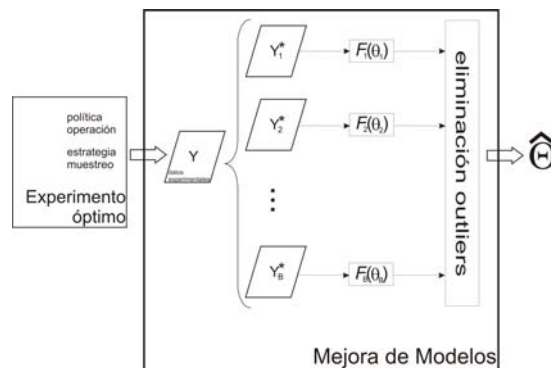
Integración de *Bootstrapping* al Diseño de Experimentos Dinámicos basado en Modelos para Optimización

La incorporación del *Bootstrapping* en el Diseño de Experimentos Dinámicos basado en Modelos para Optimización ha sido realizada con el propósito de mejorar la estimación de la incertidumbre paramétrica tanto respecto a su dimensión como a su topología mediante la estimación empírica de las *fdp* de los parámetros de los modelos de tendencias.

Una vez que los datos experimentales son obtenidos, son generadas B muestras aleatorias con reemplazo para realizar la estimación de parámetros según:

$$F = \min_{\theta_i \in \Theta_i} \left(tr \left[\left(\left(Y_n^{\text{exp}} - Y(\theta_{l-p}, \theta_l) \right)^2 \right)^T \square \left(Y_n^{\text{exp}} - Y(\theta_{l-p}, \theta_l) \right)^2 \right] \right) \quad (\text{II.5})$$

Obtenidas las B mejores estimaciones de los parámetros del modelo, las F_i son filtradas utilizando según el criterio de conservar aquellas dentro del rango: $[F_{5\text{percentilo}} - F_{95\text{percentilo}}]$. Las *fdp* empíricas de los parámetros re-estimados en cada iteración son construidas utilizando las parametrizaciones asociadas a las F_i seleccionadas. La estrategia utilizada está indicada en la siguiente ilustración:



Bibliografía

- Banks H.T., Holm K., Robbins D. (2010); Standard error computations for uncertainty quantification in inverse problems: Asymptotic theory vs. bootstrapping; *Mathematical and Computer Modelling* **52**, 1610 -1625.
- Del Barrio, E. Cuestas-Albertos, J.A.; Maltrán, C. (2002). Asymptotic stability of the bootstrap sample mean. *Stochastic Processes and their Applications*, **97**(2); 289-306; 2002.
- Chernick, M.R. (1999); *Bootstrap Methods, A Practitioner's Guide*, Wiley, New York.
- Cantelli, F. P. (1933). Sulla determinazione empirica delle leggi di probabilita. *Giorn. Ist. Ital. Attuari* **4**, 221-424.
- Efron B. and Tibshirani R. (1986), *Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy*; *Statist. Sci.* **1**(1), 54-75.
- De Martini D. y Rapallo F. (2008); On multivariate smoothed bootstrap consistency ; *Journal of Statistical Planning and Inference*; **138**, 1828 – 1835.
- Diccio, T.; Efron, B. (1996). Bootstrap Confidence Intervals. *Statistical Science*; **11**; 189-228.
- Glivenko, V. (1933). Sulla determinazione empirica della legge di probabilita. *Giorn. Ist. Ital. Attuari* **4**, 92-99.
- Henderson A. (2005); The bootstrap: A technique for data-driven statistics. Using computer-intensive analyses to explore experimental data; *Clinica Chimica Acta*; **359**; 1 -26;.
- Kochanski G. (2005). Brute Force as a Statistical Tool; <http://kochanski.org/gpk>.
- Martin M. (2007). Bootstrap hypothesis testing for some common statistical problems: A critical evaluation of size and power properties. *Computational Statistics & Data Analysis*, **51**; 6321-6342.

Anexo III

Secuencia 1:

MRGIALFVAASLIVEGTAESSICSDFGNEFCRNAECEVVPGAEDDFVCKCPRDNMYFNAAEKQCEYKDT
CKTRECSYGRCVESNPSKGSVCERSDDLTLQCKIKNDYATDCRNRRGGTAKLR TDGF IGATCDCGEWGAM
NKTTRNCVPTTCLRPDLTCKDLCEKNLLQRDSRCCQGWNTANCSAAPPADSYCSPGSPKGPDPGQCKNACR
TKEAGFVCKHGRSTDKAYECTCPSGSTVAEDGITCKSISYTVSCTVEQKQTCRPTEDCRVQKGTVLCEC
PWNQHLVGDTCISDCVDKKCHEEFMDCGVYMNRSYCPWKS RKP GPNV NINERLLNEYYYTVSFTPNIS
FDS DHCKRYEDRVLGAIRTSIGKEVFKVEILNCTQDIKARLIAEKPLSKYVLRK LQACEHPIGEWCMYP
KLLIKKNSATEIEEENL CDSLLKNQEAAYKGQNKCVKVDNLFWFQCADGYTTTYEMTRGLRRSVCKAGV
SCNENEQLECANKGQICVYENKANCQCPPDTKPGEIGCIERTT CNPKEIQECQDKKLECVYKNHKAECK
CPDDECSR

Secuencia 2:

ATG CGT GGT ATT GCG CTG TTT GTG GCG GCG GTG AGC CTG ATT GTG GAA GGT ACC GCG GAA
AGC AGC ATT TGC AGC GAT TTT GGT AAT GAA TTT TGC CGT AAT GCG GAA TGC GAA GTG GTG
CCG GGT GCG GAA GAT GAT TTT GTG TGC AAA TGC CCG CGT GAT AAT ATG TAT TTT AAT GCG
GCG GAA AAA CAG TGC GAA TAT AAA GAT ACC TGC AAA ACC CGT GAA TGC AGC TAT GGT CGT
TGC GTG GAA AGC AAT CCG AGC AAA GGT AGC TGC GTG TGC GAA CGT AGC GAT GAT CTG ACC
CTG CAG TGC AAA ATT AAA AAT GAT TAT GCG ACC GAT TGC CGT AAT CGT GGT GGT ACC GCG
AAA CTG CGT ACC GAT GGT TTT ATT GGT GCG ACC TGC GAT TGC GGT GAA TGG GGT GCG ATG
AAT AAA ACC ACC CGT AAT TGC GTG CCG ACC ACC TGC CTG CGT CCG GAT CTG ACC TGC AAA
GAT CTG TGC GAA AAA AAT CTG CTG CAG CGT GAT AGC CGT TGC TGC CAG GGT TGG AAT ACC
GCG AAT TGC AGC GCG GCG CCG CCG GCG GAT AGC TAT TGC AGC CCG GGT AGC CCG AAA GGT
CCG GAT GGT CAG TGC AAA AAT GCG TGC CGT ACC AAA GAA GCG GGT TTT GTG TGC AAA CAT
GGT TGC CGT AGC ACC GAT AAA GCG TAT GAA TGC ACC TGC CCG AGC GGT AGC ACC GTG GCG
GAA GAT GGT ATT ACC TGC AAA AGC ATT AGC TAT ACC GTG AGC TGC ACC GTG GAA CAG AAA
CAG ACC TGC CGT CCG ACC GAA GAT TGC CGT GTG CAG AAA GGT ACC GTG CTG TGC GAA TGC
CCG TGG AAT CAG CAT CTG GTG GGT GAT ACC TGC ATT AGC GAT TGC GTG GAT AAA AAA TGC
CAT GAA GAA TTT ATG GAT TGC GGT GTG TAT ATG AAT CGT CAG AGC TGC TAT TGC CCG TGG
AAA AGC CGT AAA CCG GGT CCG AAT GTG AAT ATT AAT GAA CGT CTG CTG AAT GAA TAT TAT
TAT ACC GTG AGC TTT ACC CCG AAT ATT AGC TTT GAT AGC GAT CAT TGC AAA CGT TAT GAA
GAT CGT GTG CTG GGT GCG ATT CGT ACC AGC ATT GGT AAA GAA GTG TTT AAA GTG GAA ATT
CTG AAT TGC ACC CAG GAT ATT AAA GCG CGT CTG ATT GCG GAA AAA CCG CTG AGC AAA TAT
GTG CTG CGT AAA CTG CAG GCG TGC GAA CAT CCG ATT GGT GAA TGG TGC ATG ATG TAT CCG
AAA CTG CTG ATT AAA AAA AAT AGC GCG ACC GAA ATT GAA GAA GAA AAT CTG TGC GAT AGC
CTG CTG AAA AAT CAG GAA GCG GCG TAT AAA GGT CAG AAT AAA TGC GTG AAA GTG GAT AAT
CTG TTT TGG TTT CAG TGC GCG GAT GGT TAT ACC ACC ACC TAT GAA ATG ACC CGT GGT CGT
CTG CGT CGT AGC GTG TGC AAA GCG GGT GTG AGC TGC AAT GAA AAT GAA CAG CTG GAA TGC
GCG AAT AAA GGT CAG ATT TGC GTG TAT GAA AAT GGT AAA GCG AAT TGC CAG TGC CCG CCG
GAT ACC AAA CCG GGT GAA ATT GGT TGC ATT GAA CGT ACC ACC TGC AAT CCG AAA GAA ATT
CAG GAA TGC CAG GAT AAA AAA CTG GAA TGC GTG TAT AAA AAT CAT AAA GCG GAA TGC AAA
TGC CCG GAT GAT CAT GAA TGC AGC CGT

Secuencia 3:

ATG CGT GGT ATT GCA CTG TTC GTT GCG GCA GTT TCT CTG ATC GTA GAA GGT ACC GCG GAA
TCT TCT ATC TGC AGC GAC TTC GGT AAC GAA TTC TGC CGT AAC GCT GAA TGT GAG GTA GTT
CCA GGT GAG GAA GAC GAC TTT GTT TGC AAA TGC CCG CGT GAC AAC ATG TAC TTC AAC GCT
GCC GAA AAA CAG TGC GAA TAC AAA GAC ACC TGC AAA ACC CGT GAA TGT TCC TAC GGT CGT
TGC GTT GAA TCC AAC CCG AGC AAA GGT TCT TGC GTT TGT GAA CGT TCC GAC GAC CTG ACC
CTG CAG TGC AAA ATC AAA AAC GAC TAC GCT ACT GAC TGT CGT AAC CGT GGT GGT ACC GCG
AAA CTG CGT ACG GAC GGT TTC ATT GGT GCA ACC TGC GAT TGT GGT GAA TGG GGT GCG ATG
AAC AAA ACG ACT CGT AAC TGT GTT CCG ACC ACT TGC CTG CGT CCG GAT CTG ACC TGT AAA
GAC CTC TGC GAA AAA AAC CTG CTG CAG CGT GAT TCT CGT TGT TGT CAG GGT TGG AAC ACC
GCA AAC TGT TCT GCA GCG CCA CCG GCG GAT TCT TAC TGT TCT CCG GGT TCT CCG AAA GGT
CCG GAC GGT CAG TGC AAA AAC GCT TGC CGT ACG AAA GAA GCT GGT TTC GTG TGC AAA CAC
GGT TGC CGT TCT ACG GAC AAA GCT TAC GAG TGC ACC TGT CCG TCT GGT TCT ACG GTA GCA
GAA GAC GGT ATC ACC TGC AAA TCT ATC TCT TAC ACG GTA TCT TGC ACC GTC GAA CAG AAA
CAG ACT TGT CGT CCG ACT GAA GAC TGT CGT GTT CAG AAA GGT ACC GTT CTG TGT GAA TGT
CCG TGG AAC CAG CAC CTG GTA GGT GAT ACC TGT ATC AGC GAC TGC GTC GAC AAA AAA TGC
CAC GAG GAG TTC ATG GAC TGC GGT GTT TAC ATG AAC CGT CAG AGC TGC TAC TGT CCG TGG
AAA TCT CGT AAA CCG GGT CCA AAC GTG AAC ATC AAC GAA CGT CTG CTG AAC GAG TAC TAC
TAC ACC GTG TCT TTC ACG CCG AAC ATC TCT TTC GAC AGC GAC CAC TGC AAA CGT TAC GAA
GAT CGT GTA CTG GGT GCG ATC CGT ACC TCT ATC GGC AAA GAA GTG TTC AAA GTG GAA ATC
CTG AAC TGC ACC CAG GAT ATC AAA GCC CGT CTG ATC GCT GAA AAA CCG CTG AGC AAA TAC
GTG CTG CGT AAA CTG CAG GCT TGT GAA CAC CCG ATT GGT GAA TGG TGC ATG ATG TAC CCG
AAA CTG CTG ATC AAA AAA AAC TCC GCG ACC GAA ATC GAG GAA GAA AAC CTG TGC GAC TCT
CTG CTG AAA AAC CAG GAA GCC GCC TAC AAA GGC CAG AAC AAA TGC GTT AAA GTG GAC AAC
CTC TTC TGG TTC CAG TGC GCA GAC GGT TAC ACT ACT ACT TAC GAG ATG ACT CGT GGT CGT
CTG CGT CGT TCT GTT TGC AAA GCG GGT GTT TCT TGC AAC GAG AAC GAG CAG CTG GAA TGT
GCG AAC AAA GGT CAG ATC TGC GTC TAC GAG AAC GGT AAA GCT AAC TGT CAG TGT CCA CCG
GAC ACT AAA CCG GGT GAA ATC GGT TGT ATC GAA CGT ACC ACC TGC AAC CCG AAA GAA ATC
CAG GAA TGC CAG GAC AAA AAA CTG GAG TGC GTG TAC AAA AAC CAC AAA GCC GAG TGC AAA
TGC CCG GAT GAT CAC GAA TGT TCT CGT TAA

Secuencia 4:

CAT ATG CGT GGT ATT GCG CTG TTT GTG GCG GCG GTT TCT CTG ATC GTA GAA GGT ACC GCG
GAA TCT TCT ATC TGC AGC GAC TTC GGT AAC **GAA TTC** TGC CGT AAC GCT GAA TGT GAG GTA
GTT CCA GGT GAG GAA GAC GAC TTT GTT TGC AAA TGC CCG CGT GAC AAC ATG TAC TTC AAC
GCT GCC GAA AAA CAG TGC GAA TAC AAA GAC ACC TGC AAA ACC CGT GAA TGT TCC TAC GGT
CGT TGC GTT GAA TCC AAC CCG AGC AAA GGT TCT TGC GTT TGT GAA CGT TCC GAC GAC CTG
ACC CTG CAG TGC AAA ATC AAA AAC GAC TAC GCT ACT GAC TGT CGT AAC CGT GGT GGT ACC
GCG AAA CTG CGT ACG GAC GGT TTC ATT GGT GCA ACC TGC GAT TGT GGT GAA TGG GGT GCG
ATG AAC AAA ACG ACT CGT AAC TGT GTT CCG ACC ACT TGC CTG CGT CCG GAT CTG ACC TGT
AAA GAC CTC TGC GAA AAA AAC CTG CTG CAG CGT GAT TCT CGT TGT TGT CAG GGT TGG AAC
ACC GCA AAC TGT TCT GCA GCG CCA CCG GCG GAT TCT TAC TGT TCT CCG GGT TCT CCG AAA
GGT CCG GAC GGT CAG TGC AAA AAC GCT TGC CGT ACG AAA GAA GCT GGT TTC GTG TGC AAA
CAC GGT TGC CGT TCT ACG GAC AAA GCG TAC GAG TGC ACC TGT CCG TCT GGT TCT ACG GTA
GCA GAA GAC GGT ATC ACC TGC AAA TCT ATC TCT TAC ACG GTA TCT TGC ACC GTC GAA CAG
AAA CAG ACT TGT CGT CCG ACT GAA GAC TGT CGT GTT CAG AAA GGT ACC GTT CTG TGT GAA
TGT CCG TGG AAC CAG CAC CTG GTA GGT GAT ACC TGT ATC AGC GAC TGC GTC GAC AAA AAA
TGC CAC GAG GAG TTC ATG GAC TGC GGT GTT TAC ATG AAC CGT CAG AGC TGC TAC TGT CCG
TGG AAA TCT CGT AAA CCG GGT CCA AAC GTG AAC ATC AAC GAA CGT CTG CTG AAC GAG TAC
TAC TAC ACC GTG TCT TTC ACG CCG AAC ATC TCT TTC GAC AGC GAC CAC TGC AAA CGT TAC
GAA GAT CGT GTA CTG GGT GCG ATC CGT ACC TCT ATC GGC AAA GAA GTG TTC AAA GTG GAA
ATC CTG AAC TGC ACC CAG GAT ATC AAA GCC CGT CTG ATC GCT GAA AAA CCG CTG AGC AAA
TAC GTG CTG CGT AAA CTG CAG GCT TGT GAA CAC CCG ATT GGT GAA TGG TGC ATG ATG TAC
CCG AAA CTG CTG ATC AAA AAA AAC TCC GCG ACC GAA ATC GAG GAA GAA AAC CTG TGC GAC
TCT CTG CTG AAA AAC CAG GAA GCC GCC TAC AAA GGC CAG AAC AAA TGC GTT AAA GTG GAC
AAC CTC TTC TGG TTC CAG TGC GCA GAC GGT TAC ACT ACT ACT TAC GAG ATG ACT CGT GGT
CGT CTG CGT CGT TCT GTT TGC AAA GCG GGT GTT TCT TGC AAC GAG AAC GAG CAG CTG GAA
TGT GCG AAC AAA GGT CAG ATC TGC GTC TAC GAG AAC GGT AAA GCT AAC TGT CAG TGT CCA
CCG GAC ACT AAA CCG GGT GAA ATC GGT TGT ATC GAA CGT ACC ACC TGC AAC CCG AAA GAA
ATC CAG GAA TGC CAG GAC AAA AAA CTG GAG TGC GTG TAC AAA AAC CAC AAA GCC GAG TGC
AAA TGC CCG GAT GAT CAC GAA TGT TCT CGT *AAG CTT*

Secuencia 5

NNNcNNNNNNNNNNNNNNNNNNNTcactaaagGGGAATTGTGAGCGGAatAACAAATTC
CCCTCTAGAAATAatNNNtNTttAACTTTAAGAAGGAGATATACATATGCGTGG
TATTGcGCTGTTTGTGGCGGCGGTTTCTCTGATCGTAGAAGGTACCGCGGAA
TCTTCTATCTGCAGCGACTTCGGTAACNAATTCTGCCGTAACGCTGAATGTG
AGGTAGTTCCAGGTGAGGAAGACGACTTTGTTTGCAAATGCCCGCGTGACA
ACATGTA CTTCACGCTGCCGAAAAACAGTGCGAATACAAAGACACCTGCA
AAACCCGTGAATGTTCCCTACGGTTCGTTGCGTTGAATCCAACCCGAGCAAAG
GTTCTTGCGTTTGTGAACGTTCCGACGACCTGACCCTGCAGTGCAAAATCAA
AAACGACTACGCTACTGACTGTCGTAACCGTGGTGGTACCGCGAAACTGCG
TACGGACGGTTTCATTGGTGCAACCTGCGATTGTGGTGAATGGGGTGCAT
GAACAAAACGACTCGTAACTGTGTTCCGACCACTTGCCTGCGTCCGGATCTG
ACCTGTAAAGACCTCTGCGAAAAAACCTGCTGCAGCGTGATTCTCGTTGTT
GTCAGGGTTGGAACACCGCAAACCTGTTCTGCAGCGCCACCGGCGGATTCTT
ACTGTTCTCCGGGTTCTCCGAAAGGTCCGGACGGTCAGTGCAAAAACGCTT
GCCGTACGAAAGAAGCTGGTTTTCGTGTGCAAACACGGTTGCCGTTCTACGG
ACAAAGCGTACGAGTGCACCTGTCCGTCTGGTTCTACNGtAGCAGAAGACGG
TATCACCTGCAAATCTATCTCTTACACGGTATCTTGCACCGTCGAACAGAAA
CaGACTTGTTCGTCGACTGAAGACTGTCGTGTTTCAGAAAGGTACCGTTCTGT
GTGAATGTNCGTGGAANCNGCACctgGNNaGGTGATACCTGTATCAGCgACT
GCGTCNACAAAAATGN

Secuencia 6

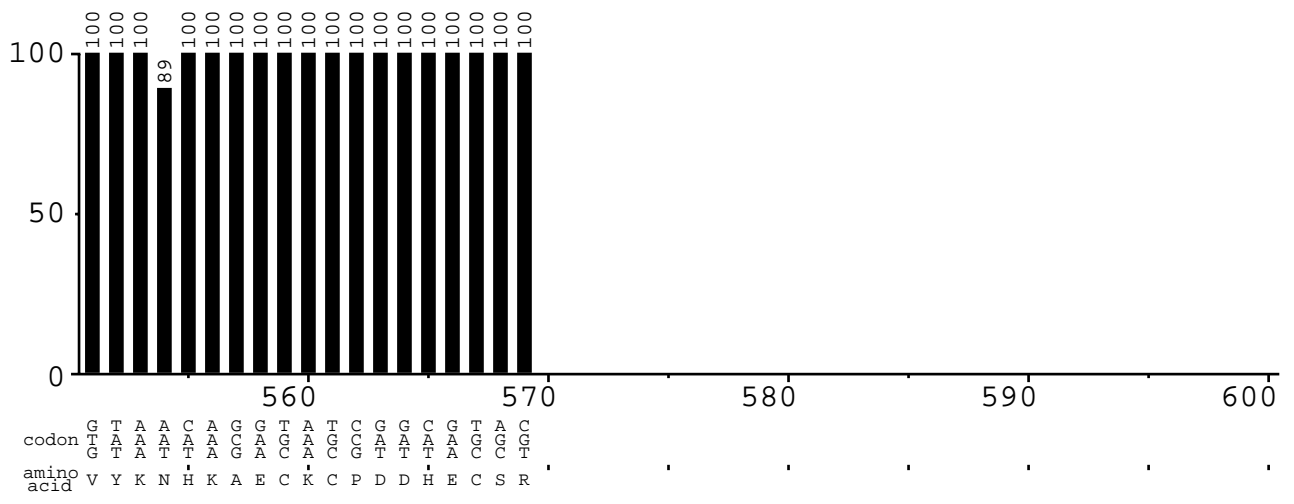
NNNNNNNNNttcttttNcgNgNctttgtagCagCCGGaTCTCagtGGTGGTGGTGGTGG
TGCTCGAGtgCggtcgCaAGCTTACGAGAACATTCGTGATCATCCGGGCaTTTGC
ACTCGGCTTTGTGGTTTTTGTACACGCACTCCAGTTTTTTTGTCTGGCATTCC
TGGATTTCTTTTCGGGTTGCAGGTGGTACGTTTCGATAACAACCGATTTACCCCG
GTTTAGTGTCCGGTGGACACTGACAGTTAGCTTTACCGTTCTCGTAGACGCA
GATCTGACCTTTGTTTCGCACATTCCAGCTGCTCGTTCTCGTTGCAAGAAACA
CCCGCTTTGCAAACAGAACGACGCAGACGACCACGAGTCATCTCGTAAGTA
GTAGTGTAACCGTCTGCGCACTGGAACCAGAAGAGGTTGTCCaCTTTAAcgca
TTTGTCTGGCCTTTGTAGGCGGCTTCCTGGTTTTTTCAGCAgAgAgtcgCACAG
GTTTTCTTCCTCGATTTcGtgcggAGTTTTTTTTGATCAGCAGTTTCGGGtACA
TCATGCACCATTCACCAATCGGGTGttCacaAgcctGcAGtTTacgCagCACGTATTT
GcTCagCNGGTTTTTTCagCgatcagACGGNCTTTTGATATCCNtGGGTGCagttCANG
AtttccAcNTTGAACACTTCTTTGNCCgAtagANgtACGGATCGcACCCagtACACgA
TCTTcgtAACgNTTGCAGtGGTCgCTGtcgaNgagatGtTCNGNgtGNAAgAcaCNGtGT
AgtAgtACTCgtNcagcaNAcGTTcgNTGAtGTNcACgNttNGNANCCGGNTNACNAga
NtNCCNCgNAcaGgAgCANctctGAcGGTTCNtGTNAAcaCCgcaNteaTGNANctCNteg
NNcaNTTTTTGtgaNgcagNcgetGANNcaNGtatNNNCCTaCCANGtGcTGNtNcacNNg
acaTcaCNcaGNacNGNtaccNTTntNNaNaNaNNgTCTccaNtc

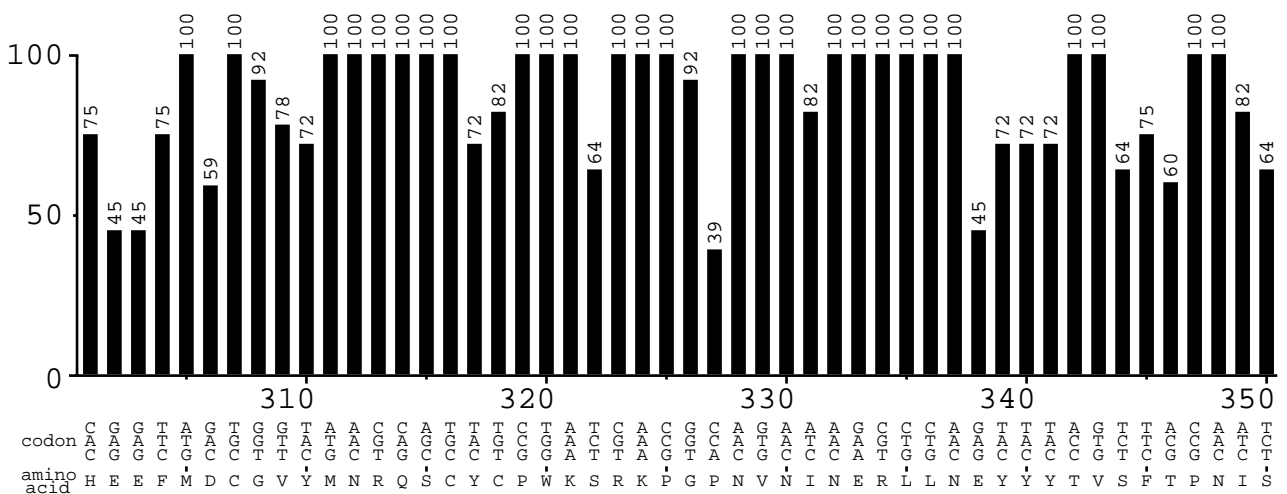
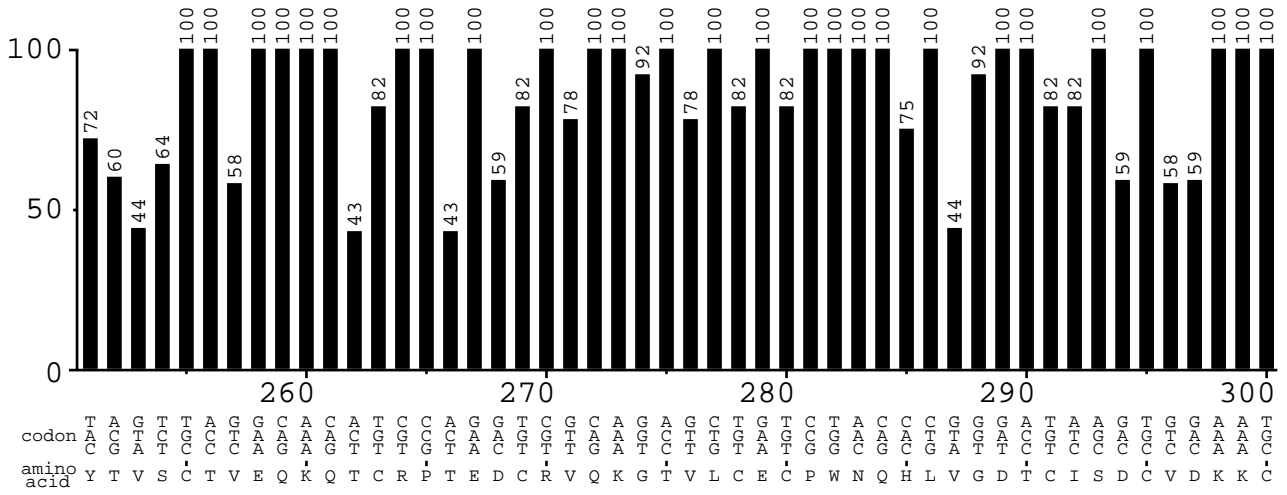
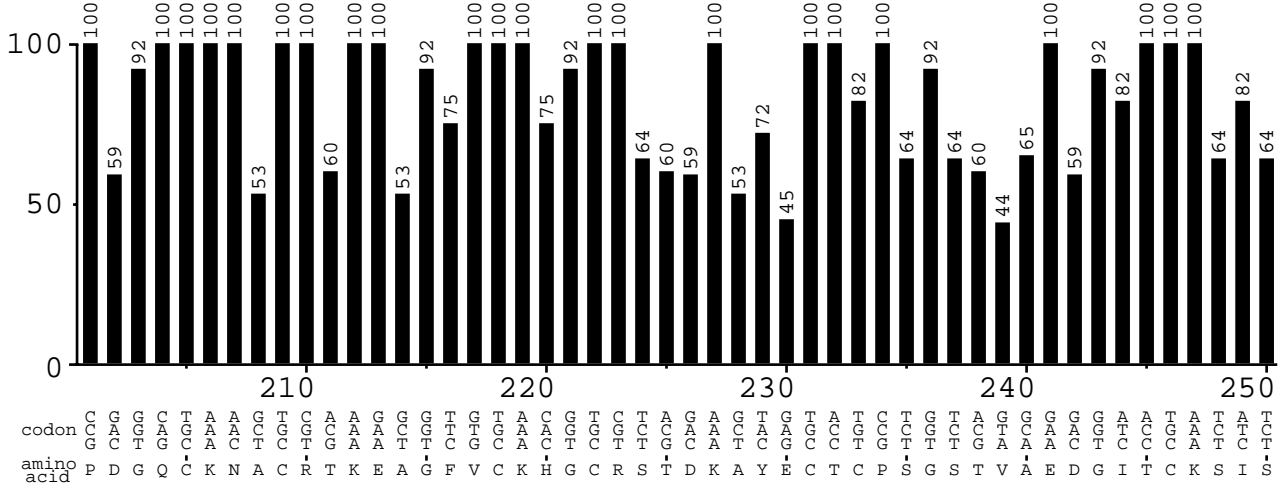
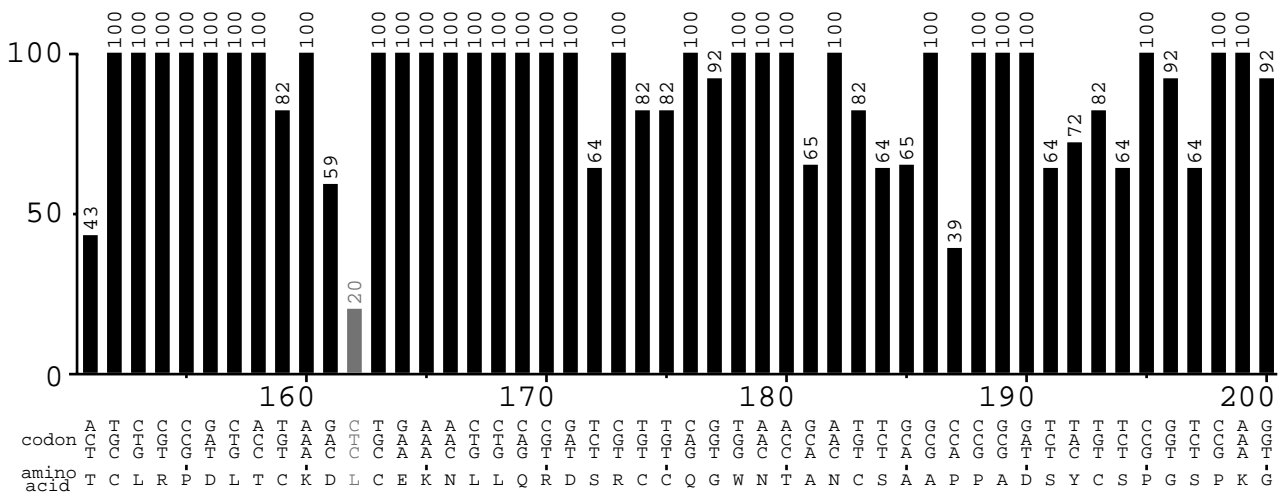
Secuencia 7:

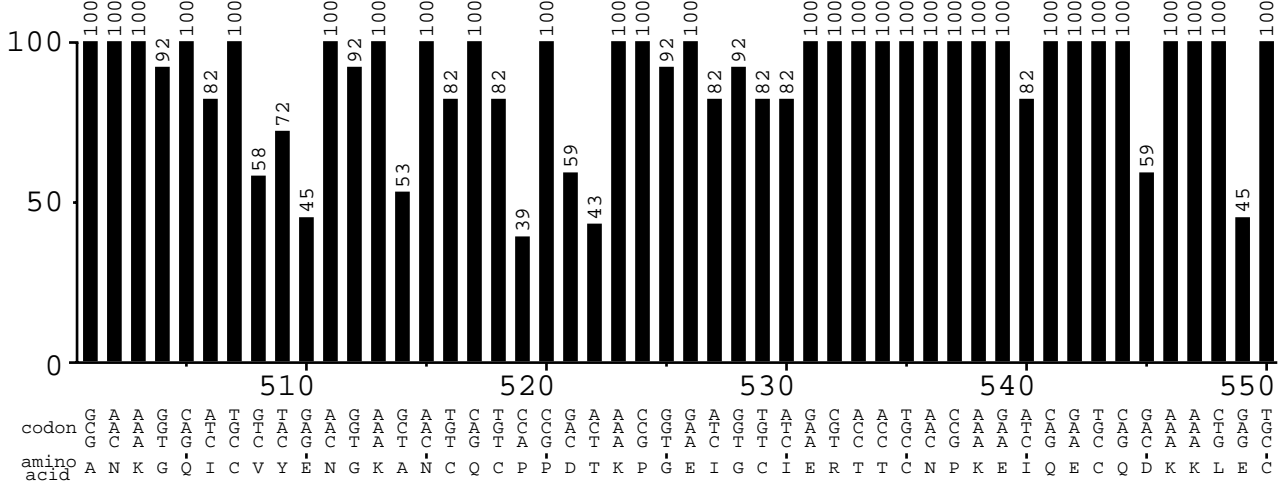
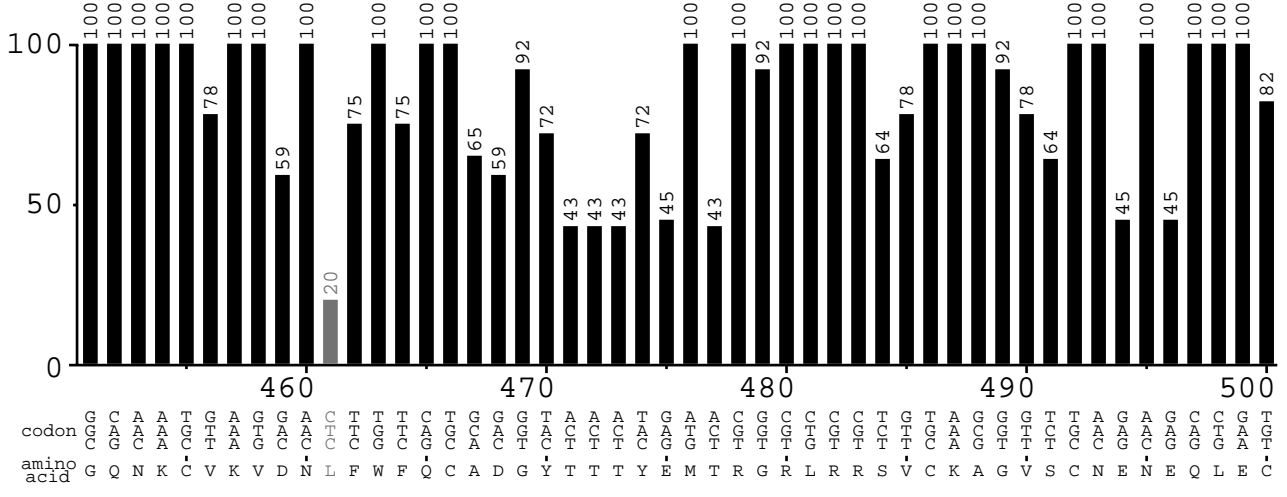
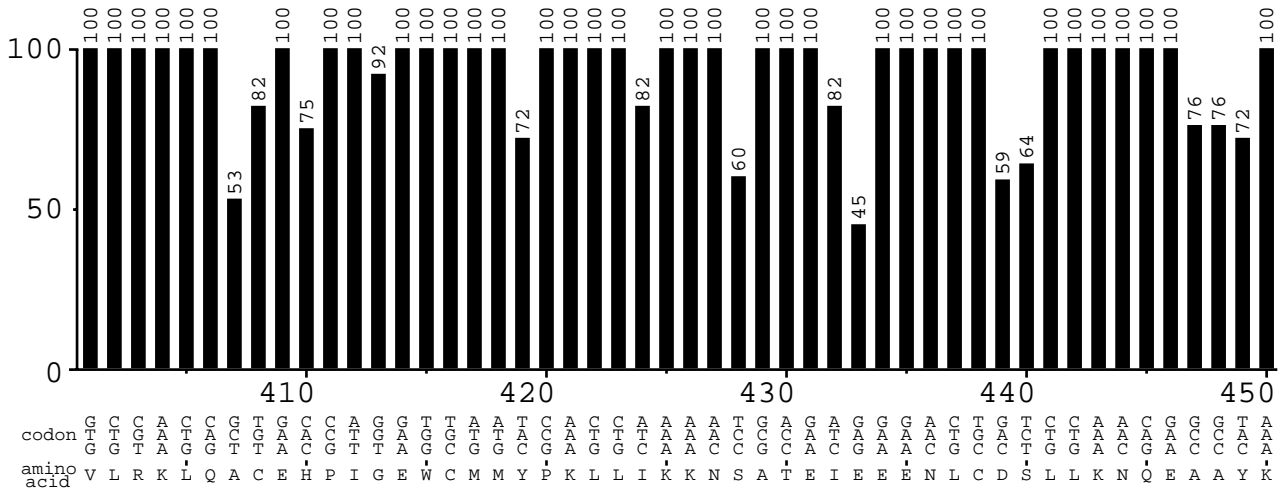
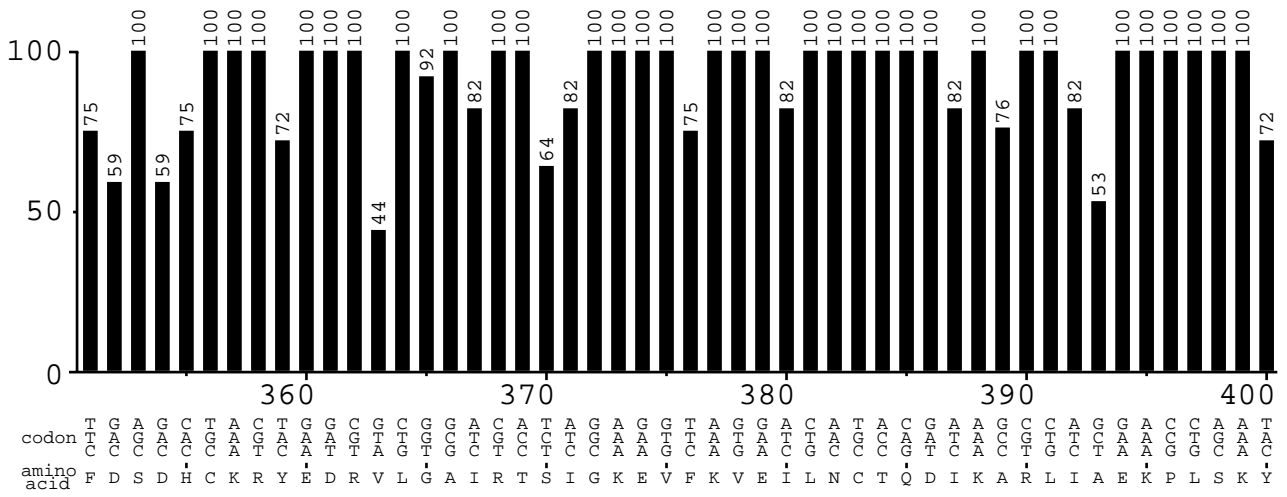
ANNGNNGNNTCTTNNTGTTCT
CCGGGTTCTCCGAAAGGTCCGGACGGTCAGTNNNNNNNNNNcTTGcCGTACGA
AAGAAGCTGGTTTTTCGTGTGCAAACACgGNTGCCGTTCTACGGACAAAGCGT
ACGAGTGCACCTGTCCGTCTGGTTCTACGGTAGCAGAAGACGGTATCACCT
GCAAATCTATCTCTTACACGGTATCTTGCACCGTCGAACAGAAACAGACTTG
TCGTCCGACTGAAGACTGTCGTGTTTCAGAAAGGTACCGTTCTGTGTGAATGT
CCGTGGAACCAGCACCTGGTAGGTGATACCTGTATCAGCGACTGCGTCGAC
AAAAAATGCCACGAGGAGTTCATGGACTGCGGTGTTTACATGAACCGTCAG
AGCTGCTACTGTCCGTGGAAATCTCGTAAACCGGGTCCAAACGTGAACATC
AACGAACGTCTGCTGAACGAGTACTACTACACCGTGTCTTTCACGCCGAAC
ATCTCTTTCGACAGCGACCACTGCAAACGTTACGAAGATCGTGTACTGGGTG
CGATCCGTACCTCTATCGGCAAAGAAGTGTTCAAAGTGGAATCCTGAACT
GCACCCAGGATATCAAAGCCCGTCTGATCGCTGAAAAACCGCTGAGCAAAT
ACGTGCTGCGTAAACTGCAGGCTTGTGAACACCCGATTGGTGAATGGTGCA
TGATGTACCCGAAACTGCTGATCAAAAAAACTCCGCGACCGAAATCGAGG
AAGAAAACCTGTGCGACTCTCTGCTGAAAAACCAGGAAGCCGCCTACNaagg
cCAGAACAAATGCGTTAAAGTGGACAACCTCTTCTGGTTCCNGTGCGCAGAC
gGTTACACTACTACTTACGAGATGACTCGTGGTCGTCTGCGTCGTTCTGTTTG
CAAAGCGGGTGTTCCTTGCAACGAGAACGAGCAGCTGGAATGTGCGAACAA
ANGGTCANATCTGCGTCTACNANAACGGTAAAGCTNACTGTCAGTGNCCNN
CNGNANNCTAAANCNGGGNGAAATCNGTNGNNTCNANNGNACCNNCNNN
NNCNNNNAANNNNNCNNNNNNNN

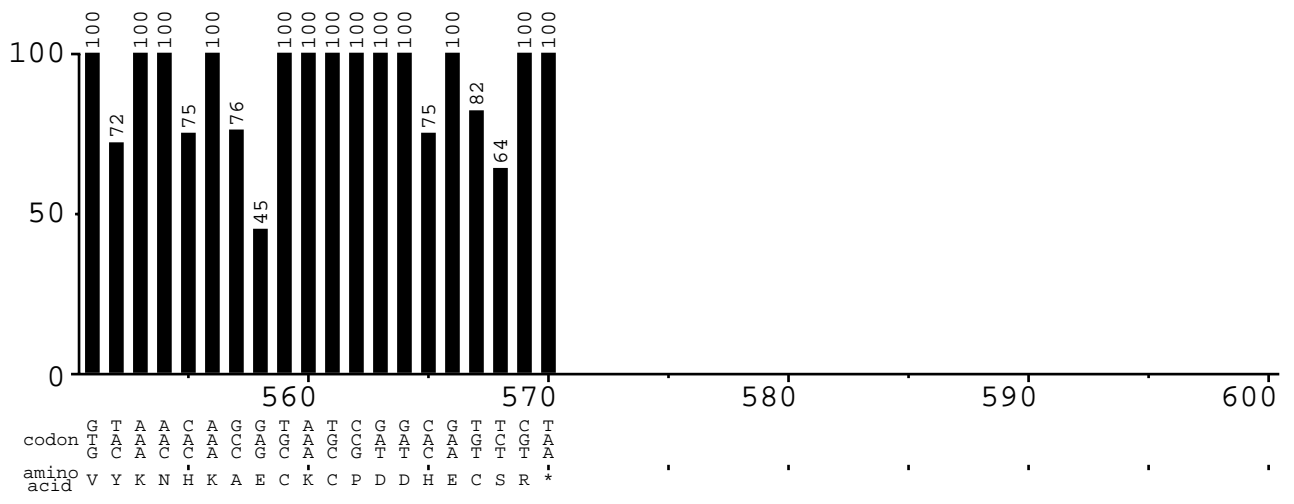
Secuencia 8:

NNNcNNNNNNNNNNNNNNNNNNNTcactaaagGGGAATTGTGAGCGGAatAACAAATTC
CCCTCTAGAAATAatNNNtNTttAACTTTAAGAAGGAGATATACATATGCGTGG
TATTGCGCTGTTTGTGGCGGCGGTTTCTCTGATCGTAGAAGGTACCGCGGAA
TCTTCTATCTGCAGCGACTTCGGTAACNAATTCTGCCGTAACGCTGAATGTG
AGGTAGTTCCAGGTGAGGAAGACGACTTTGTTTGCAAATGCCCGCGTGACA
ACATGTACTIONTCAACGCTGCCGAAAAACAGTGCGAATACAAAGACACCTGCA
AAACCCGTGAATGTTCCCTACGGTTCGTTGCGTTGAATCCAACCCGAGCAAAG
GTTCTTTCGTTTTGTGAACGTTCCGACGACCTGACCCTGCAGTGCAAAATCAA
AAACGACTACGCTACTGACTGTCGTAACCGTGGTGGTACCGCGAAACTGCG
TACGGACGGTTTTATTGGTGCAACCTGCGATTGTGGTGAATGGGGTTCGAT
GAACAAAACGACTCGTAACTGTGTTCCGACCACTTGCCTGCGTCCGGATCTG
ACCTGTAAAGACCTCTGCGAAAAAACCTGCTGCAGCGTGATTCTCTCGTTGT
TGTCAGGGTTGGAACACCGCAAACCTGTTCTGCAGCGCCACCGGCGGATTCT
TACTGTTCTCCGGGTTCTCCGAAAGGTCCGACGGTCAGTGCAAAAACGCTT
GCCGTACGAAAGAAGCTGGTTTTCTGTGTGCAAACACGGTTGCCGTTCTACGG
ACAAAGCGTACGAGTGCACCTGTCCGTCTGGTTCTACGGTAGCAGAAGACG
GTATCACCTGCAAATCTATCTCTTACACGGTATCTTGCACCGTCGAACAGAA
ACAGACTTGTGCTCCGACTGAAGACTGTCGTGTTTCAGAAAGGTACCGTTCTG
TGTGAATGTCCGTGGAACCAGCACCTGGTAGGTGATACCTGTATCAGCGAC
TGCGTCGACAAAAAATGCCACGAGGAGTTCATGGACTGCGGTGTTTACATG
AACCGTCAGAGCTGCTACTGTCCGTGGAAATCTCGTAAACCGGGTCCAAAC
GTGAACATCAACGAACGTCTGCTGAACGAGTACTACTACACCGTGTCTTTCA
CGCCGAACATCTCTTTCGACAGCGACCACTGCAAACGTTACGAAGATCGTG
TACTGGGTGCGATCCGTACCTCTATCGGCAAAGAAGTGTTCAAAGTGGA
TCCTGAACTGCACCCAGGATATCAAAGCCCGTCTGATCGCTGAAAAACCGC
TGAGCAAATACGTGCTGCGTAAACTGCAGGCTTGTGAACACCCGATTGGTG
AATGGTGCATGATGTACCCGAAACTGCTGATCAAAAAAACTCCGCGACCG
AAATCGAGGAAGAAAACCTGTGCGACTCTCTGCTGAAAAACCAGGAAGCCG
CCTACNaagcCAGAACAATGCGTTAAAGtGGACAACCTCTTCTGGTTCCAGT
GCGCAGACGGTTACTACTACTTACGAGATGACTCGTGGTCGTCTGCGTCG
TTCTGTTTGCAAAGCGGGTGTTCCTTGAACGAGAACGAGCAGCTGGAATGT
GCGAACAAAGGTCAGATCTGCGTCTACGAGAACGGTAAAGCTAACTGTCAG
TGTCCACCGGACACTAAACCGGGTGAAATCGGTTGTATCGAACGTACCACC
TGCAACCCGAAAGAAATCCAGGAATGCCAGGACAAAAAACTGGAGTGCCT
GTACAAAAACCACAAAGCCGAGTGCAAAAtGCCCGGATGATCACGAATGTTC
TCGTAAGCTtGcgaccGcaCTCGAGCACCACCACCACCactGAGAtCCGGetGct
aacaagNcNcgNaaaagaaNNNNNNNNN









rBM95_opt_letol0bien

threshold1: 20% (grey)

threshold2: 10% (red)

sequence derived from

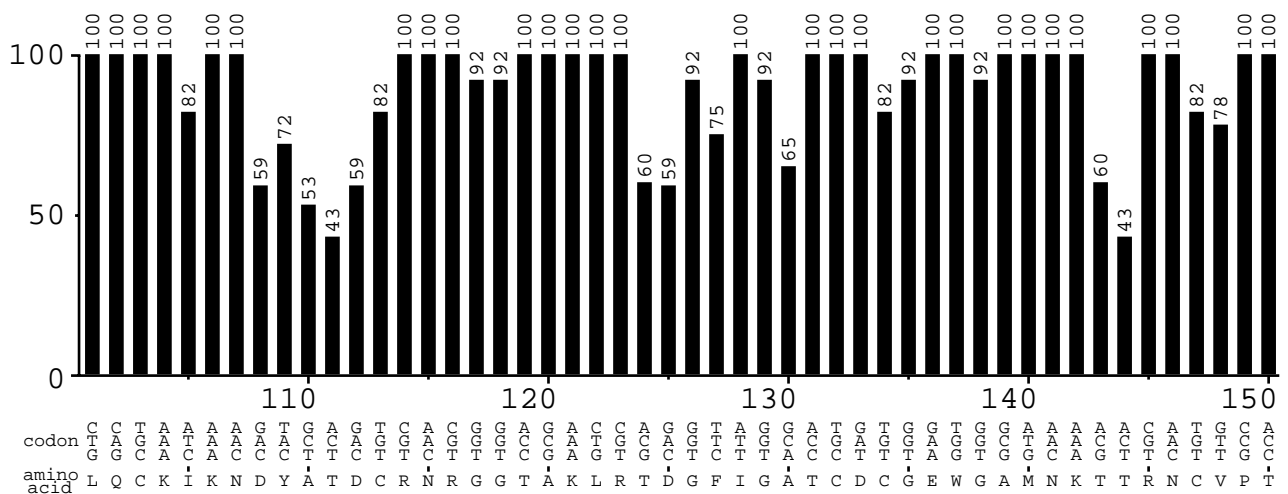
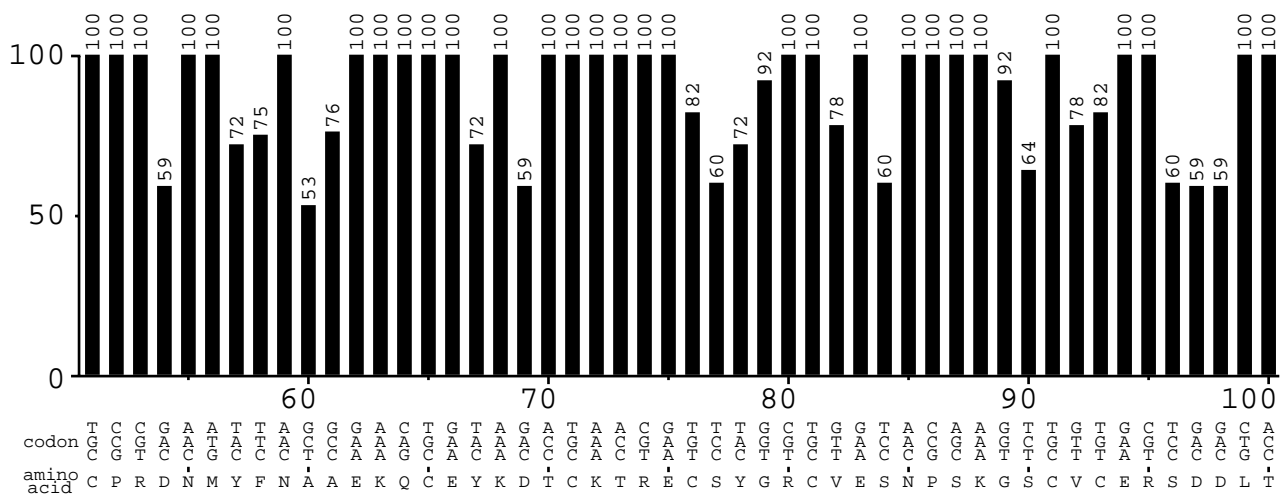
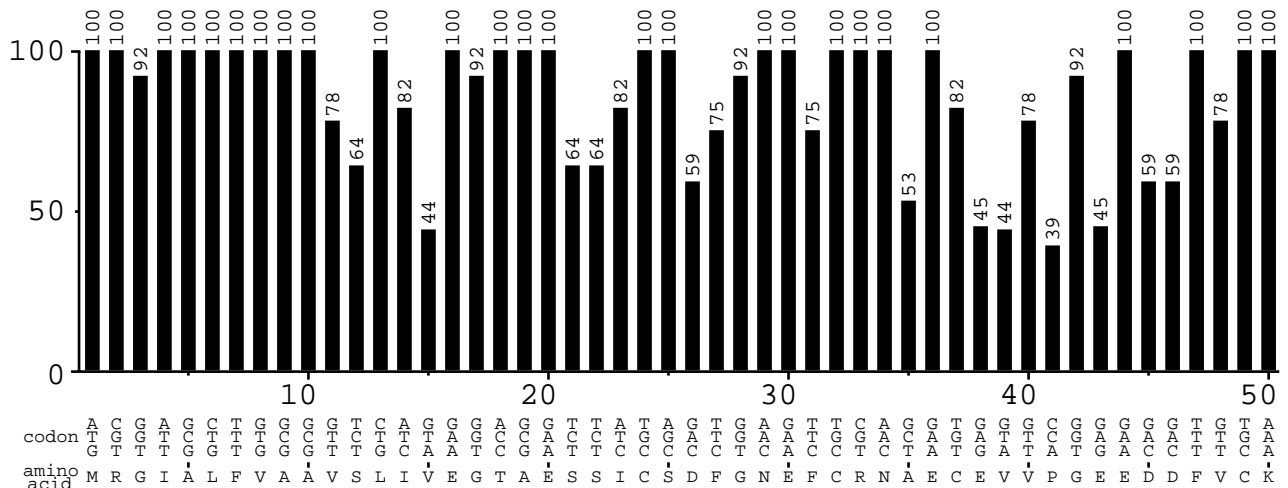
Boophilus microplus

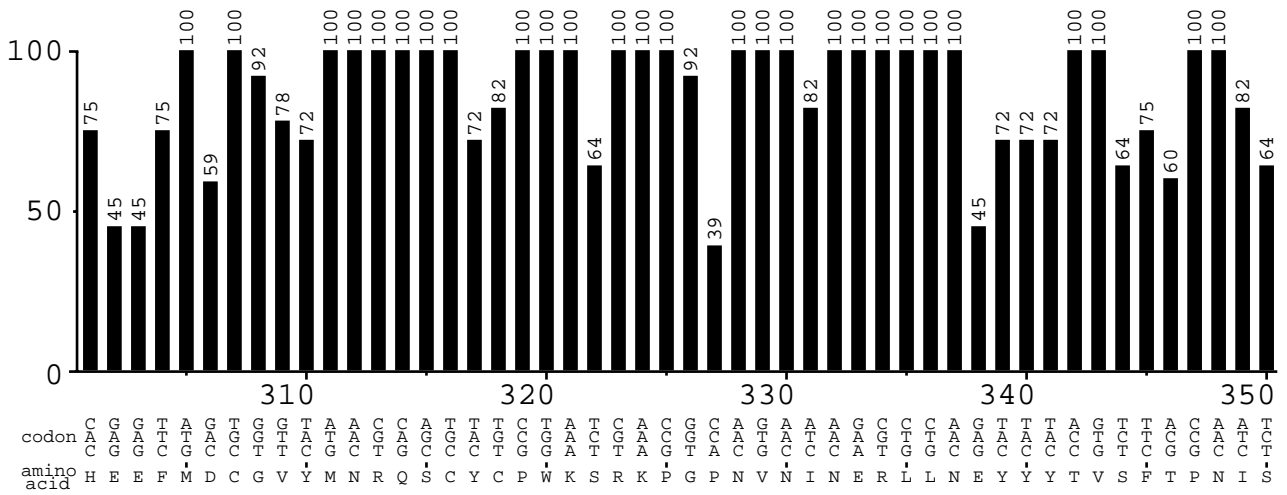
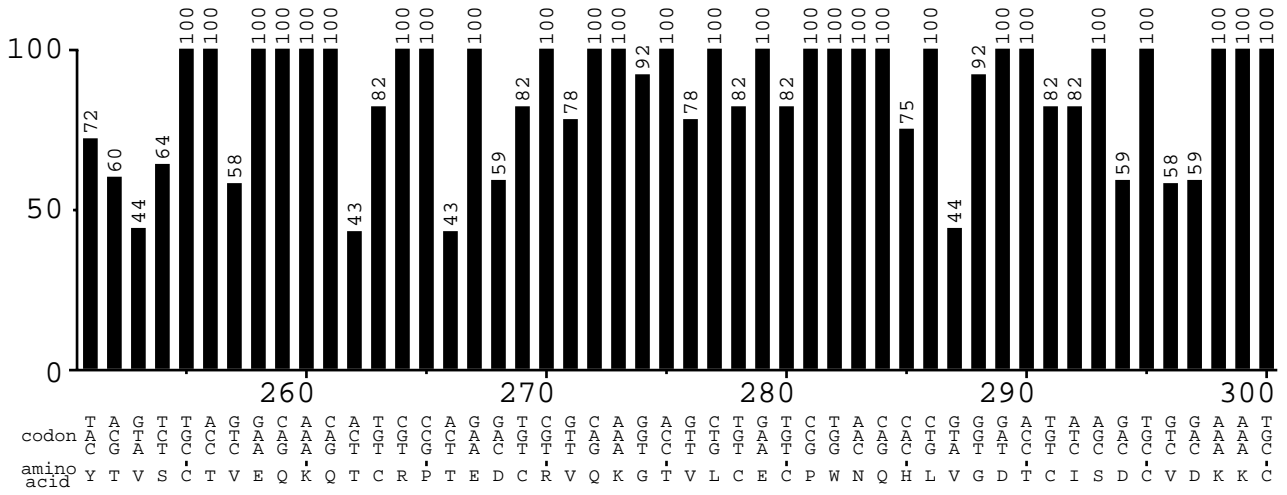
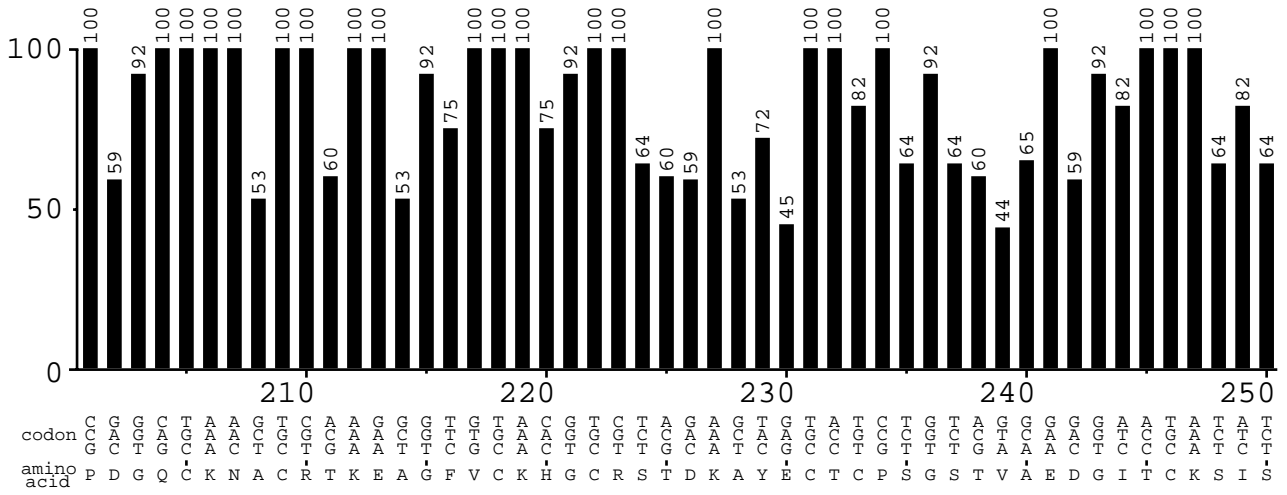
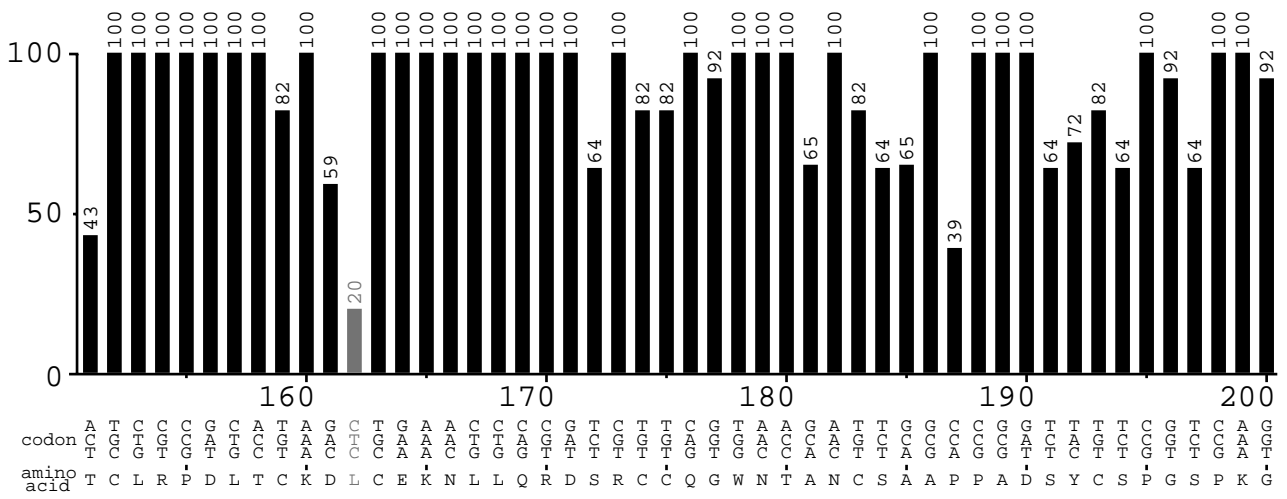
www.gcuu.de

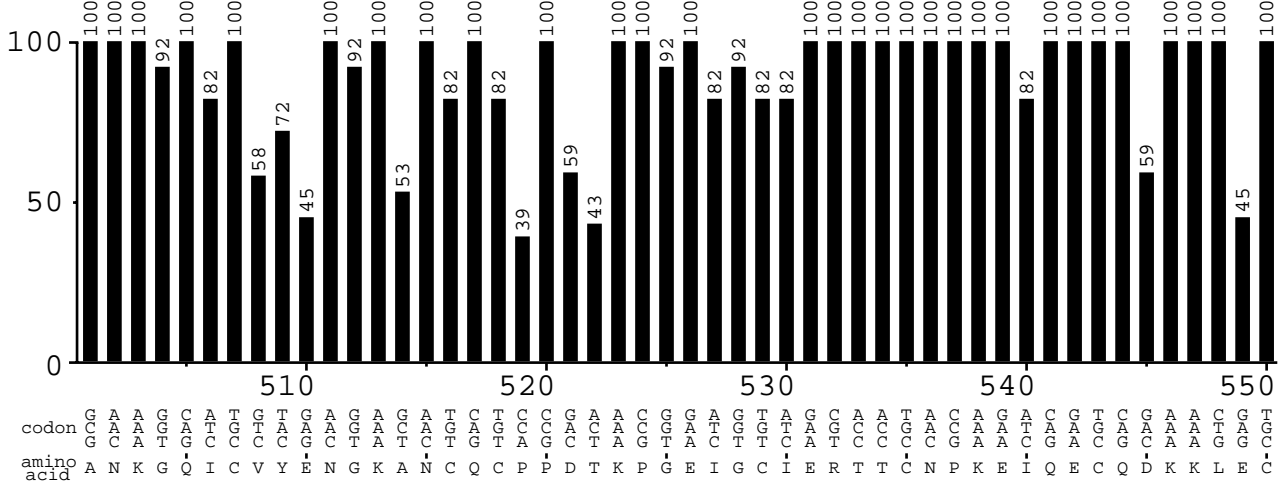
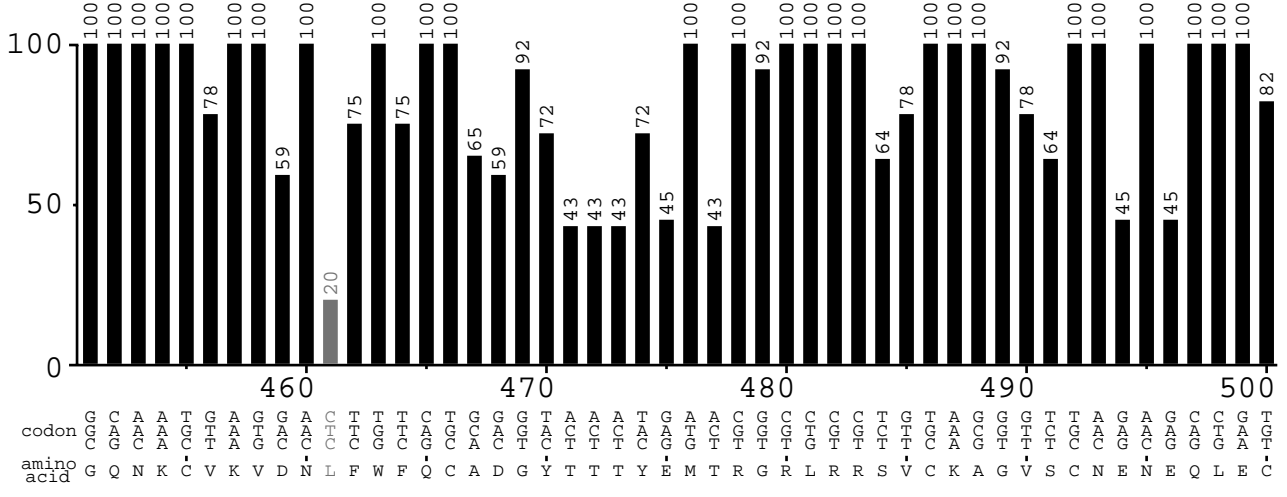
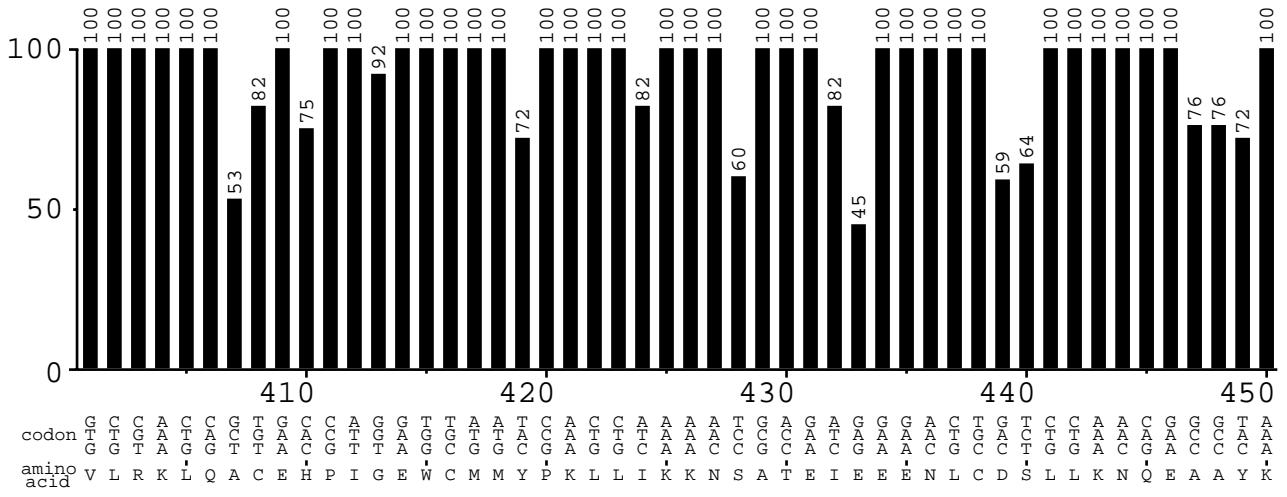
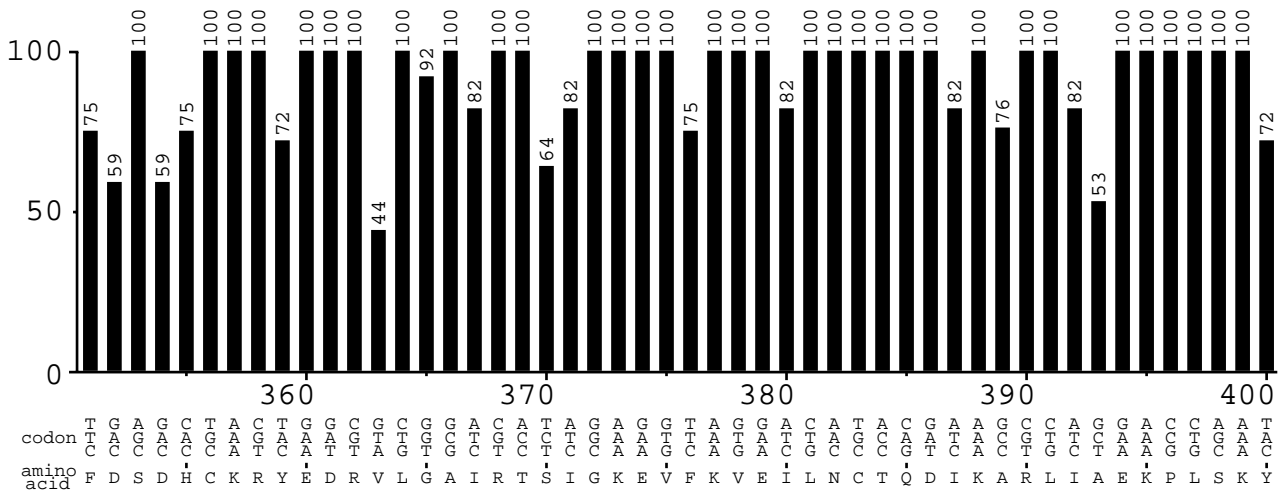
created: 12.05.2007

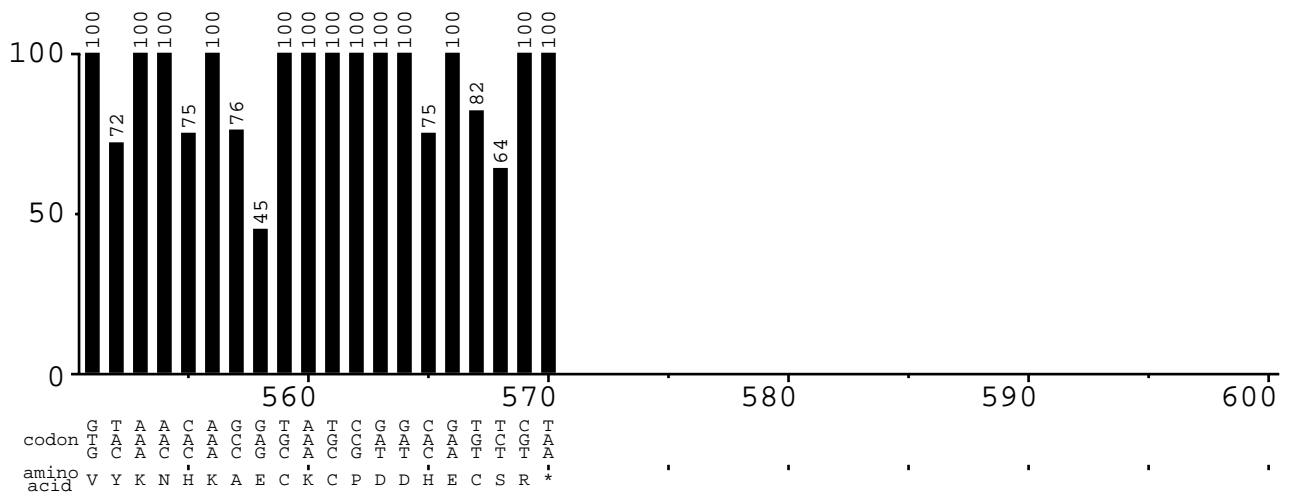
Codontable:
Escherichia_coli

Ordinate (y-axis): relative adaptiveness









Al i neami ento 1

Al i neami ento 1:

Bm95 protei n [Boophi l us mi cropl us]
Length=569

Score = 615 bits (1585), Expect(2) = 8e-180
Identi ties = 280/285 (98%), Posi tives = 280/285 (98%), Gaps = 0/285 (0%)
Frame = +2

Query	101	MRGI ALFVAAVSLI VEGTAESSI CSDFGNXFCRNAECEVVPGEEDDFVCKCPRDNMYFNA	280
		MRGI ALFVAAVSLI VEGTAESSI CSDFGN FCRNAECEVVPG EDDFVCKCPRDNMYFNA	
Sbj ct	1	MRGI ALFVAAVSLI VEGTAESSI CSDFGNEFCRNAECEVPGAEDDFVCKCPRDNMYFNA	60
Query	281	AEKQCEYKDTCKTRECSYGRCVESNPSKGSCVCERSDDLTLCKI KNDYATDCRNRRGGTA	460
		AEKQCEYKDTCKTRECSYGRCVESNPSKGSCVCERSDDLTLCKI KNDYATDCRNRRGGTA	
Sbj ct	61	AEKQCEYKDTCKTRECSYGRCVESNPSKGSCVCERSDDLTLCKI KNDYATDCRNRRGGTA	120
Query	461	KLRTDGF I GATCDCGEWGAMNKTRNCVPTTCLRPDLTCKDLCEKNLLORDSRCCQGWNT	640
		KLRTDGF I GATCDCGEWGAMNKTRNCVPTTCLRPDLTCKDLCEKNLLORDSRCCQGWNT	
Sbj ct	121	KLRTDGF I GATCDCGEWGAMNKTRNCVPTTCLRPDLTCKDLCEKNLLORDSRCCQGWNT	180
Query	641	ANCSAAPPADSYCSPGSPKGPDGQCKNACRTKEAGFVCKHGRSTDKAYECTCPSGSTVA	820
		ANCSAAPPADSYCSPGSPKGPDGQCKNACRTKEAGFVCKHGRSTDKAYECTCPSGSTVA	
Sbj ct	181	ANCSAAPPADSYCSPGSPKGPDGQCKNACRTKEAGFVCKHGRSTDKAYECTCPSGSTVA	240
Query	821	EDGI TCKSI SYTVSCTVEQKQTCRPTEDCRVQKGTVLCECXWXXH	955
		EDGI TCKSI SYTVSCTVEQKQTCRPTEDCRVQKGTVLCEC W H	
Sbj ct	241	EDGI TCKSI SYTVSCTVEQKQTCRPTEDCRVQKGTVLCECPWNQH	285

Al i neami ento 2

Al i neami ento 2:

Bm95 protei n [Boophi l us mi cropl us]
Length=569

Score = 395 bits (1014), Expect = 1e-112
Identi ties = 201/244 (82%), Posi ti ves = 210/244 (86%), Gaps = 5/244 (2%)
Frame = -3

Query	809	VXI XERLXEYyytvsxtxni sstattaxv--TKI VYWVRSVXLSXKEVFXVEI XNCTXG	636
		V I ER L EYyyTVS T NIS + +++ +R+ KEVF VEI NCT	
Sbj ct	329	VNI NERLLNEYyyTVSFTPNI SFDS DHCKRYEDRVLGAI RTSI --GKEVFKVEI LNCTQD	386
Query	635	YQKXV*SLKNXLSKYVLRKLOACEHPI GEWCMMYPKLLI KKNSATEI EEENLCDSLLKNQ	456
		+ + + K LSKYVLRKLOACEHPI GEWCMMYPKLLI KKNSATEI EEENLCDSLLKNQ	
Sbj ct	387	I KARLI AEK-PLSKYVLRKLOACEHPI GEWCMMYPKLLI KKNSATEI EEENLCDSLLKNQ	445
Query	455	EAAYKGQNKCVKVDNLFWFQCADGYTTTYEMTRGRLRRSVCKAGVSCNENEQLECANCKGQ	276
		EAAYKGQNKCVKVDNLFWFQCADGYTTTYEMTRGRLRRSVCKAGVSCNENEQLECANCKGQ	
Sbj ct	446	EAAYKGQNKCVKVDNLFWFQCADGYTTTYEMTRGRLRRSVCKAGVSCNENEQLECANCKGQ	505
Query	275	I CVYENGKANCQCPPDTKPGEI GCI ERTTCNPKEI QECQDKKLECVYKNHKAECKCPDDH	96
		I CVYENGKANCQCPPDTKPGEI GCI ERTTCNPKEI QECQDKKLECVYKNHKAECKCPDDH	
Sbj ct	506	I CVYENGKANCQCPPDTKPGEI GCI ERTTCNPKEI QECQDKKLECVYKNHKAECKCPDDH	565
Query	95	ECSR 84	
		ECSR	
Sbj ct	566	ECSR 569	

Al i neami ento 4:

>gb|AAD38381.2|AF150891_1 Bm95 protei n [Rhi pi cephal us mi cropl us]
 Length=569

Score = 1232 bits (3188), Expect = 0.0
 Identi ties = 566/569 (99%), Posi ti ves = 566/569 (99%), Gaps = 0/569 (0%)
 Frame = +2

Query	101	M RGI ALFVAAVSLI VEGTAESSI CSDFGNXFCR NAECEVVPGEEDDFVCKPRDNMYFNA	280
Sbj ct	1	M RGI ALFVAAVSLI VEGTAESSI CSDFGN FCR NAECEVVPG EDDFVCKPRDNMYFNA	60
Query	281	A EKQCEYKDTCKTRECSYGR CVESNPSKGSCVCERSDDLTLCKI KNDYATDCRN RGGTA	460
Sbj ct	61	A EKQCEYKDTCKTRECSYGR CVESNPSKGSCVCERSDDLTLCKI KNDYATDCRN RGGTA	120
Query	461	K LR TDGFI GATCDCGEWGAMNKTRNCVPTTCLRPDLTCKDLCEKNLLORDSRCCQGWNT	640
Sbj ct	121	K LR TDGFI GATCDCGEWGAMNKTRNCVPTTCLRPDLTCKDLCEKNLLORDSRCCQGWNT	180
Query	641	A NCSAAPPADSYCSPGSPKPGDQCKNACRTKEAGFVCKHGC RSTDKAYECTCPSGSTVA	820
Sbj ct	181	A NCSAAPPADSYCSPGSPKPGDQCKNACRTKEAGFVCKHGC RSTDKAYECTCPSGSTVA	240
Query	821	E DGI TCKSI SYTVSCTVEQKQTCRPTEDCRVQKGTVLCECPWNQHLVGDTCI SDCVDKKC	1000
Sbj ct	241	E DGI TCKSI SYTVSCTVEQKQTCRPTEDCRVQKGTVLCECPWNQHLVGDTCI SDCVDKKC	300
Query	1001	H EEFMDCGVYMNROSCYCPWKS RKP GPNVNI NERLLNEYYYTVSFTPNI SFDS DHCKRYE	1180
Sbj ct	301	H EEFMDCGVYMNROSCYCPWKS RKP GPNVNI NERLLNEYYYTVSFTPNI SFDS DHCKRYE	360
Query	1181	D RVLGAI RTSI GKEVFKVEI LNCTQDI KARLI AEKPLSKYVLRKLOACEHPI GEWCMMYP	1360
Sbj ct	361	D RVLGAI RTSI GKEVFKVEI LNCTQDI KARLI AEKPLSKYVLRKLOACEHPI GEWCMMYP	420
Query	1361	K LLI KKNSATEI EEENLCDSLLKNQEAAYXGONKCVKVDNLFWFQCADGYTTTYEMTRGR	1540
Sbj ct	421	K LLI KKNSATEI EEENLCDSLLKNQEAAY GONKCVKVDNLFWFQCADGYTTTYEMTRGR	480
Query	1541	L RRSVCKAGVSCNENEQLECAN KGOI CVYENKANCQC PPDTKPGEI GCI ERTTCNPKEI	1720
Sbj ct	481	L RRSVCKAGVSCNENEQLECAN KGOI CVYENKANCQC PPDTKPGEI GCI ERTTCNPKEI	540
Query	1721	Q ECQDKKLECVYKNHKAECKPDDHECSR	1807
Sbj ct	541	Q ECQDKKLECVYKNHKAECKPDDHECSR	569

Anexo IV: Técnicas microbiológicas y de Biología Molecular

A.4.1 Medios de cultivo

La composición de los medios de cultivos utilizados es mostrada en las tablas A.4.1 y 2.

Tabla A.4.1. 1: Medio LB selectivo (Luria Bertani)

Componente	Concentración (g L ⁻¹)
Triptona de Caseína	10
Extracto de Levaduras	5
NaCl	5
Kanamicina	30 10 ⁻³

Tabla A.4.1. 2: Medio TB selectivo (Terrific Broth)

Componente	Concentración (g L ⁻¹)
Triptona de Caseína	12
Extracto de Levaduras	24
KH ₂ PO ₄	2,31
K ₂ HPO ₄	12,54
Kanamicina	30 10 ⁻³

En los casos en los que fueron utilizados medios sólidos las composiciones contaron con agar-agar 1%.

A.4.2 Transformación de ADN en *E. coli*.

Preparación de células de *E. coli* competentes.

Las cepas de *E. coli* utilizadas para la transformación con ADN fueron Top10 y BL21(DE3). Para la preparación de las células competentes se inocularon 5 ml de medio de cultivo LB con una colonia bacteriana e incubado a 37°C, hasta que alcanzó una densidad óptica de 0,6 a 600 nm. Se tomó una alícuota de 1 mL del cultivo y se mantuvo en agua/hielo durante 5 minutos. Seguidamente las células se recogieron por centrifugación a 3.000 x g, durante 5 minutos y se resuspendieron en MgCl₂ 100 mM, previamente enfriado en hielo, hasta 1/5 del volumen del cultivo y se incubaron en agua/hielo durante 20 minutos.

Se cosecharon nuevamente por centrifugación (3.000 x g, 5 minutos) y se resuspendieron en 1/2 del volumen inicial en CaCl₂ 100 mM, para ser incubadas durante 1 hora en agua/hielo.

Transformación de células bacterianas competentes.

El proceso de transformación de *E. coli* fue mediante choque térmico. A la suspensión de células competentes se le añadieron 20-50 ng de ADN plasmídico circular, o bien toda la mezcla de ligación (hasta 400 ng), y se mantuvo en agua/hielo durante 20 minutos. A continuación, la mezcla se incubó a 42°C durante 1 minuto y nuevamente se

mantuvo en agua/hielo durante 5 minutos. Tras este enfriamiento se añadió 1 ml de medio LB a temperatura ambiente y se incubó durante 1 hora a 37°C con agitación suave, para permitir la recuperación de las bacterias. Por último, las bacterias se recogieron por centrifugación a 3.000 x g, 5 minutos, se resuspendieron en 200 µL de medio fresco y se sembraron en placas Petri que contenían medio de cultivo sólido LB selectivo. Las placas se incubaron a 37°C de 12 a 16 horas.

Selección de los clones transformantes de *E. coli*.

Todos los plásmidos utilizados portan el gen de resistencia a la kanamicina, por lo que la selección de los transformantes se basó en el hecho de que sólo las células transformadas pueden crecer en un medio que contenga este antibiótico. La selección de los clones que portaban los plásmidos recombinantes se realizó inoculando clones al azar en 5 ml de medio LB-kanamicina y creciendo el cultivo 12 horas. Se extrajo el ADN plasmídico y se digirió con diversas enzimas de restricción para verificar la presencia del ADN inserto, del que también se determinó su secuencia nucleotídica.

A.4.3 Miniprep de ADN plasmídico

Una cantidad adecuada de células es cosechada (3.000g, 5 min) y resuspendida en 250 µL de solución adecuada y 250 µL de solución de lisis. La mezcla es incubada durante 1-5 min. Luego se agregan 350 µL de solución de neutralización y sobrenadante es transferido a una columna de purificación de ADN. Se centrifuga y el filtrado es descartado. La columna es lavada con solución de lavado y posterior centrifugación y el ADN es eluido con agua libre de nucleasas.

A.4.4 Digestión del ADN con Endonucleasas de Restricción.

Las digestiones del ADN con endonucleasas de restricción se llevaron a cabo según las condiciones de buffer de reacción y temperatura especificadas por el fabricante. El volumen total de reacción fue de 100 µL □ y el tiempo de incubación fue de 2 horas para la digestión de plásmidos a 37°C.

Las digestiones dobles se realizaron digiriendo con las dos enzimas de restricción a la vez cuando el tampón de reacción fue común para las dos enzimas.

Tabla A.4.4. 1: Mezcla de reacción digestión

Reactivo	Volumen (µL)
Buffer Tango (2x)	20
Plásmido	45
NdeI	3,5
HindIII	4,5
Agua	27

A.4.5 Electroforesis de ácidos nucleicos en geles de agarosa.

La separación de los fragmentos de ADN en función de su tamaño, se llevó a cabo en geles de agarosa 1% (p/v). Las electroforesis se desarrollaron en cubetas horizontales a voltaje constante (100 V) y como buffer se utilizó TAE 1x (buffer TAE 50x: Tris-HCl 2 M, pH 7,2, EDTA 50 mM, pH 8, ácido acético glacial 57,1 ml y agua bidestilada hasta 1 L). Las muestras se prepararon mezclando el ADN con buffer de carga de alta densidad, compuesto por azul de bromofenol 0,25% (p/v) y glicerol 30% (v/v). Para la visualización del ADN se añadió bromuro de etidio (0,2 ng/ml) al gel de agarosa y se le hizo incidir luz ultravioleta de longitud de onda de 340 nm a través de un transiluminador.

A.4.6 Fosforilación T4 kinasa

La reacción de fosforilación se realizó siguiendo las especificaciones del fabricante. Brevemente, 16 µL de solución de fragmentos de ADN purificados se les agregó 2 µL buffer T4 ligasa y 2 µL de T4 kinasa. La mezcla fue incubada a 37°C durante 20 minutos y luego la enzima fue inactivada por incubación a 75°C durante 10 minutos.

A.4.7 Electroforesis de proteínas en geles de poliacrilamida (SDS- PAGE).

Las muestras de proteína fueron sometidas a electroforesis en geles de poliacrilamida según el método descrito por Laemmli (1970). Las electroforesis se llevaron a cabo en el sistema Mini-Protean II Cell (Bio-Rad). La separación de las proteínas fue desarrollada sobre geles discontinuos en los que la concentración de acrilamida fue de 10 % de reticulación (pH 8.8) para la zona de resolución y de 4% para la de *stacking* (pH 6.8).

Las proteínas antes de ser cargadas en el gel se hirvieron durante 10 minutos en buffer de carga de Laemmli [Tris-HCl 62,5 mM, pH 6,8, SDS 2% (p/v), glicerol 10%, □□Mercaptoetanol 1% (v/v), azul de bromofenol 0,002% (p/v)]. La electroforesis se desarrolló en buffer de electroforesis de proteínas SDS-PAGE 1x [SDS-PAGE 10x: Tris-HCl 250 mM, glicina 144 mM, SDS 10%], a amperaje constante.

**Tabla A.4.7 1: Marcador de Peso Molecular
Prestained SDS PAGE Standars Broad Range**

Marca: Bio-Rad.

Catalogo 161-0318, Control 310005521

Protein	Calibrated MW on Tris-HCl gel	Calibrated MW on Bis-Tris gel	Calibrated MW on Tris Acetate gel
Myosin	198,844	195,785	197,603
b-galactosidase	115,7	104,02	110,271
Bovine serum albumin	96,743	59,84	72,274
Ovoalbumin	53,541	41,675	44,495
Carbonic anhydrase	37,134	27,855	31,6
Sybean trypsin inhibitor	29,134	21,168	
Lysozyme	19,54	15,223	
Aprotinin	6,905	6,501	

Tinción de proteínas en los geles de poliacrilamida.

Rutinariamente, las proteínas separadas mediante electroforesis se visualizaron tiñéndolas con azul Coomassie R-250 al 0,2% (p/v) en solución fijadora [metanol al 50% (v/v) y ácido acético al 10% (v/v)], según el método descrito por Fairbanks y col. (1971).

Los geles se destiñeron sumergiéndolos en una solución de metanol al 40% (v/v) y ácido acético al 10% (v/v)

A.4.8 Purificación de rBm95 con perlas de agarosa con quelación de Ni²⁺

La purificación de la proteína rBm95 con cola de poli-Hys ha sido realizada con perlas de agarosa siguiendo las especificaciones del fabricante.

A.4.9 Evaluación de estabilidad de construcción plasmídica pET-24b-bm95 en cepas de *E. coli* transformadas.

Con esta experiencia será evaluada la estabilidad de la construcción genética pET-24b-bm95 en la población bacteriana correspondiente a la cepa de *E. coli* BL21 transformada.

Básicamente, la técnica consiste en el cultivo de la cepa bajo estudio en un medio no selectivo (sin la presencia de antibiótico) para que un recuento de microorganismos viables en medios de cultivo agarizados suplementados y no suplementados con el antibiótico de selección sea realizado a diferentes tiempos.

El porcentaje de estabilidad de la construcción es calculado como la relación entre las UFC/mL en las placas con y sin antibiótico.

Procedimiento de Recuento en placa en superficie

La cepa *E. coli* BL21 transformadas con la construcción la construcción genética pET-24b-*bm95* fueron cultivadas en medio TB glucosado (10 g L^{-1}) no selectivo.

Se realizaron diluciones seriadas de muestras obtenidas de los cultivos a dos (2) tiempos de cultivo: 1-previa inducción; 2- cuatro (4) horas luego de la inducción de la proteína recombinante.

Se realizaron diluciones seriadas 1/10 de las muestras y de cada dilución se tomaron 0,1 mL y se dispersaron sobre la superficie de una placa de Petri con medio Luria Bertani agarizado el cual contenía o no kanamicina a una concentración de 30 mg.L^{-1} .

Anexo V: Técnicas Operatorias para la determinación de variables de estado del cultivo en biorreactor BioFlo 110.

A.5.1 Determinación de concentración celular

La concentración celular de los cultivos ha sido determinada utilizando el método gravimétrico de residuo seco. Brevemente, un volumen adecuado de cultivo es centrifugado a 3000 g, 5 minutos para cosechar las células suspendidas en un tubo previamente tarado.

Luego el *pellet* de células es resuspendido en solución fisiológica (NaCl 0.85%) y las células son nuevamente cosechadas. El sobrenadante es descartado y la muestra es colocada en estufa a 105 °C hasta peso constante.

La concentración celular es calculada utilizando la ecuación:

$$C_{\text{celular}} \left[\text{g} \cdot \text{L}^{-1} \right] = \frac{M_2 - M_1}{V_{\text{cosechado}}}$$

M_1 [g]: *tara*

M_2 [g]: *masa celular + tara*

A.5.2 Determinación de glucosa por kit enzimático

Este método involucra la reacción de oxidación de la glucosa catalizada por la enzima glucosa oxidasa (GOD) la cual produce ácido glucónico y peróxido de hidrógeno. Luego, el peróxido de hidrógeno formado reacciona ulteriormente oxidando al reactivo cromogénico (4-aminofenazona/Fenol) gracias a la acción enzimática de una peroxidasa (POD) lo que produce quinonimida cuyo color es rojo.

A las muestras se les realiza una dilución adecuada ya que el kit utilizado presenta un standard cuya concentración es 1 g L⁻¹ y el rango de concentraciones en los que la ley de Beer tiene validez es hasta 4.5 g L⁻¹.

La técnica para la determinación de glucosa es resumida en la tabla A.5.10.1.

Tabla A.5.2. 1: Procedimiento para determinación de glucosa

	Blanco	Standard	Muestra
Muestra	-	-	15 µL
Standard	-	15 µL	-
Agua	15 µL		
Rvo trabajo	1,5 mL	1,5 mL	1,5 mL

Mezclar e incubar a 37°C durante 15 minutos y leer a 505 nm.

A.5.3 Determinación de proteínas por el método de Lowry

Esta técnica se lleva a cabo en dos etapas. En la primera, las proteínas son tratadas con una solución alcalina de cobre; y en la segunda se adiciona el reactivo de Folin Ciocalteau, siendo su constituyente activo la mezcla ácida tungstica-molibdica. Las proteínas provocan la reducción de los ácidos fosfomolibdico y fosfotungstico con pérdida de 1, 2 ó 3 átomos de oxígeno del tungstato y/o molibdato, produciendo un complejo que tiene un color azul característico.

El cobre, al permanecer en la estructura peptídica, facilita la transferencia de electrones desde la vecindad de los grupos funcionales aminoacídicos, a la mezcla cromogénica ácida, incrementando así la sensibilidad de la proteína.

La producción de color por los péptidos es aproximadamente igual a la suma de cromogenicidad de los aminoácidos que los componen (dependiendo de la secuencia, longitud de la cadena y la exposición de los grupos funcionales) y de la proporcionada por las uniones peptídicas.

La tirosina y el triptofano dan color incluso en ausencia de cobre, la cistina, cisteína e histidina son también cromogénicos, pero en menor grado, pero el resto de la molécula proteica no produce la reacción coloreada a menos que existan los iones cobre.

La reacción se realiza a pH 10 debido a que a esta concentración de hidrogeniones se obtiene el máximo desarrollo de color. A ese pH la solución de Folin es solamente reactiva por un corto tiempo, razón por la cual es necesario mezclar de inmediato. La disminución de la reactividad parece ser función de la desaparición del color amarillo original del fosfomolibdato (aproximadamente 8 segundos), y es presumiblemente debido a la disociación de los aniones fosfomolibdato con formación de ácido fosfórico, que disminuye el pH, motivo por el cual la solución debe ser tamponada.

Reactivos:

- a) Solución testigo: Albúmina 100 mg%.
- b) Reactivo A: Disolver 30 gr de Na_2CO_3 anhidro y 4 gr de NaOH, en agua dest. y enrasar a 1L.
- c) Reactivo B: Solución de $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ al 2%.
- d) Reactivo C: Solución de tartrato de Na y K al 4%.

Mezcla Reactiva: Mezclar en el momento de usar 1 volumen de reactivo B con 1 volumen de reactivo C y 100 volúmenes de reactivo A .

- e) Reactivo de Folin-Ciocalteau: Solución de tungstato y molibdato de Na en medio ácido. Diluir al medio con agua destilada antes de usar.

La curva de calibrado esta en el rango [0;50] %mg.

Tabla A.5.3. 1: Procedimiento para determinación de proteínas totales

	μL
Vol. a tratar	200
Solución reactiva	1200
Rtvo de Folin	120

La solución reactiva es adicionada al volumen a tratar y la mezcla se deja reaccionar durante 10 minutos. Luego de agregado el reactivo de Folin la mezcla se deja reaccionar durante 30 minutos y leer a 660 nm.

A.5.4 Determinación de acetato por kit enzimático

Este método involucra la reacción de la acetato kinasa (Sn. 4) en presencia de ATP (Sn.2) en un medio adecuado que cataliza la formación de acetil-fosfato y 5'-difosfato de adenosina (ADP). Esa reacción es acelerada por la presencia de acetil-CoA y fosfato inorgánico por la acción de la enzima fosfotransacetilasa en presencia de coenzima A (Sn. 3). El ADP formado es reconvertido en ATP en una reacción catalizada por la piruvato kinasa en la que fosfoenolpiruvato es convertido en piruvato. Este último reacciona con NADH y se reduce a lactato en presencia de la D-lactato deshidrogenada (Sn. 5). La caída de la absorbancia de la mezcla de reacción es una señal indirecta de la presencia de acetato en la muestra.

Procedimiento

El procedimiento sugerido por el fabricante es mostrado en la tabla A.5.12.1,

Tabla A.5.4. 1: Procedimiento para la determinación de acetato

	Blanco	Muestra
Muestra	-	50 μL
Agua	1,050 mL	1,0 mL
Sn 2	250 μL	250 μL
Sn 3	10 μL	10 μL
Sn 4	10 μL	10 μL

Mezclar y leer las absorbancias de las soluciones a 340 nm (A_1) después de 2 minutos y comenzar las reacciones agregando 10 μL de la Sn 5.

Mezclar y leer las absorbancias de las soluciones a 340 nm (A_2) después de 4 minutos.

Cálculos:

Calcular ($A_1 - A_2$) y luego de eso restar la diferencia de absorbancia del blanco a las diferencias de las muestras para obtener ΔA_{acetic} (este valor debe ser mayor a 0.1 unidades de absorbancia).

La concentración de ácido acético en la muestra es calculado como:

$$c = \frac{V \cdot MW}{\epsilon \cdot d \cdot v} \cdot \Delta A_{\text{acetic}}$$

$$c = 0.2535 \cdot \Delta A_{\text{acetic}}$$