

Tesis de Licenciatura:
**Aprendizaje de Dinámicas Robustas en
Redes Regulatorias Bacterianas
Utilizando GENIE**

Alumno:

Patricio Germán Traverso
Libreta universitaria: 137/96
Correo electrónico: ptraver@dc.uba.ar

Director:

Dr. Igor Zwir

Departamento de Computación
Facultad de Ciencias Exactas y Naturales.
Universidad de Buenos Aires.

2 de septiembre de 2004

Abstract

One of the big challenges of the post genomic era is determining when, where and for how long genes are turned on or off[Bre00]. Gene expression is determined by protein-protein interactions among regulatory proteins and with RNA polymerase(s), and protein-DNA interactions of these trans-acting factors with cis-acting DNA sequences in the promoters of regulated genes[Kae03]. These interactions define complex genetic networks, which designs have motivated researchers to draw direct analogies with established techniques in electrical engineering[CEHL02], [HMC02]. As with the construction of electrical circuits, the gene circuit approach uses mathematical and computational tools in the construction and posterior analysis of a proposed network diagram. So far, the qualitative agreement between model and experiment in a series of studies depends both on the design of the network topology, which most of the times includes uncertain connections between genes, as well as on the dynamic behavior of the network, which is affected by the ambiguity inherent to the biological processes (e.g., monomer or dimer binding of promoters, enzymes having kinase and/or phosphatase activities, etc.) and the mathematical models used to represent them (e.g., Boolean or continuous models; reverse or forward algorithms)[vSWEB02]. Moreover, the number of genes considered in the networks is usually large compared to the number of the available measurements (e.g., time-point expression), thus, more than one possible model may be consistent with the subjacent data[WHA]. Finally, the data always contains a substantial amount of noise [LW01a], [LW01b], [MA99] provided by the systematic variability of the experiments, which in addition to previous problems, makes it difficult to deduce the implications of the underlying logic of genetic networks through experimental techniques alone.

We propose a methodology, termed GENIE for Gene Expression Networks Iterative Explorer, that embraces the uncertainty inherent to the biological problem and the imprecision of their underlined mathematical models by using an iterative approach. First, GENIE proposes a network topology based on DNA sequence analysis of transcription factor interactions, which, together with previous knowledge from the literature, constitute the raw material for the architecture design. Second, we transform the hypothesis provided by the network topology, by means of its possible chemical reactions and physical constraints, into a system of nonlinear ordinary differential equations, whose variables are concentrations of proteins, mRNA, etc [vDMMO00a], [MMOD02]. Rather than advocating a single, definitive model of the genetic network, we describe a variety of models that have different strengths, weaknesses and domains of applicability. Third, the network models are evaluated by testing the ability to reproduce the behavior observed *in vivo* of their subjacent non-linear models, each characterizing the time-dependent change in concentration of the components, including ki-

nase, phosphatase, and transcription activities. Fourth, the successful models are tested by considering different emergent systemic properties, such as flexibility to reproduce all possible functional patterns, and robustness to changes in parameters and initial conditions. Fifth, we revisit the original topology and iterate, developing adaptive models of genetic networks. Finally, a decision making process reveals the most realistic models, which are examined by a datamining process providing substantial insights from the modeled genetic systems.

We applied GENIE to uncover regulatory networks in the enteric bacteria *Salmonella enterica* serovar Typhimurium by focusing on the PhoP/PhoQ and PmrA/PmrB two-component systems, which govern virulence and the adaptation to low Mg^{2+} and high Fe^{3+} environments, respectively [HS95], [KLG03]. The study of the PhoP regulon constitutes a special challenge due to the multiplicity of PhoP-controlled targets, and the connectivity of the PhoP/PhoQ system with other two-component systems, such as PmrA/PmrB, transcriptional regulators, and alternative RNA polymerase sigma factors. We verified our predictions by investigating transcription and PhoP and PmrA protein binding to the identified promoters *in vivo*.

Resumen

Uno de los grandes desafíos de la era post genómica consiste en determinar *cuándo, dónde y por cuánto tiempo* los genes se encuentran activados o desactivados [Bre00]. La expresión genética es determinada por interacciones proteína-proteína entre proteínas regulatorias y RNA polimerasa(s), e interacciones de estos factores trans con secuencias de DNA en los promotores de los genes regulados [Kae03]. Estas interacciones definen complejas redes genéticas, cuyos diseños han motivado a los investigadores a realizar analogías directas con técnicas de ingeniería eléctrica [CEHL02] [HMC02]. Como en la construcción de circuitos eléctricos, el enfoque de circuitos genéticos utiliza herramientas matemáticas y computacionales para la construcción y posterior análisis del diagrama de red propuesto. Hasta el presente, el acuerdo cualitativo entre modelo y experimento en una serie de estudios depende tanto del diseño de la topología de la red, la cual la mayoría de las veces incluye conexiones desconocidas entre los genes, como del comportamiento dinámico de la misma, el cual es afectado por la ambigüedad inherente a los procesos biológicos (e.g. binding de dímeros o monómeros a promotores, enzimas con actividad kinasa y/o fosfatasa, etc.) y a los modelos matemáticos utilizados para representarla (e.g. modelos continuos o booleanos, algoritmos de ingeniería backward o feedforward, etc.) [vSWEB02]. Asimismo, el número de genes considerado en las redes es usualmente grande comparado con el número de mediciones disponibles (e.g. puntos temporales de expresión), por lo tanto, más de un modelo posible puede ser consistente con los datos subyacentes [WHA]. Finalmente, los datos siempre contienen una significativa cantidad de ruido [LW01a] [LW01b] [MA99] provisto por la variabilidad sistemática de los experimentos la cual, adicionalmente a los problemas previos, hace difícil deducir las implicaciones de la lógica subyacente de las redes genéticas a través de técnicas experimentales únicamente.

Proponemos una metodología, denominada GENIE por Gene Expression Networks Iterative Explorer (explorador iterativo de redes de expresión genética), la cual analiza la incerteza inherente al problema biológico y la imprecisión de sus modelos matemáticos subyacentes utilizando un enfoque adaptativo. En primer lugar, GENIE propone una topología de red basada en análisis de secuencias de DNA relativas a interacción de factores de transcripción la cual, junto con el conocimiento previo existente en la literatura, constituye la materia prima para el diseño de arquitecturas. En segundo lugar, se transforman las hipótesis provistas por la topología de red en base a sus posibles reacciones químicas y restricciones físicas, en un sistema de ecuaciones diferenciales ordinarias no lineales, cuyas variables son las concentraciones de proteínas, mRNA, etc. [vDMMO00a] [MMOD02]. En lugar de escoger un único, definitivo modelo de red genética, se describen una variedad de modelos con distintas fortalezas, debilidades y dominios de aplicación. En tercer lugar, en los modelos de red se evalúa la habilidad de sus

modelos no lineales subyacentes para reproducir el comportamiento observado *in vivo*, cada uno de los cuales caracteriza un cambio dependiente del tiempo en la concentración de los componentes, incluyendo actividades kinasa, fosfatasa y de transcripción. En cuarto lugar, los modelos exitosos son evaluados considerando distintas propiedades sistémicas emergentes, tales como la flexibilidad para reproducir todos los posibles patrones funcionales y la robustez con respecto a cambios en los parámetros y en las condiciones iniciales. En quinto lugar, se revisa la topología original y se itera, desarrollando modelos adaptativos de redes genéticas. Finalmente, un proceso de toma de decisiones revela los modelos más realistas, los cuales son examinados mediante un proceso de datamining proveyendo revelaciones sustanciales sobre los sistemas genéticos modelados.

Se aplicó GENIE para descubrir redes regulatorias en la bacteria *Salmonella enterica serovar Typhimurium*, focalizando el análisis en los sistemas de dos componentes PhoP/PhoQ y PmrA/PmrB, los cuales gobiernan la virulencia y la adaptación a medios de bajo Mg^{2+} y alto Fe^{3+} , respectivamente [HS95] [KLG03]. El estudio del regulón PhoP constituye un desafío especial debido a la multiplicidad de blancos controlados por PhoP, y a la conectividad del sistema PhoP/PhoQ con otros sistemas de dos componentes, tales como PmrA/PmrB, reguladores transcripcionales y factores sigma alternativos de RNA polimerasa. Las predicciones realizadas fueron verificadas investigando la transcripción y el ligamiento de las proteínas PhoP y PmrA a promotores identificados *in vivo*.

Índice general

1. Redes Genéticas	1
1.1. Introducción	1
1.2. Problema Biológico	1
1.2.1. Sistemas de Dos Componentes	1
1.2.2. Los Sistemas de Dos Componentes PhoP/PhoQ y PmrA/PmrB	2
1.3. Problema Computacional	4
1.3.1. Modelado de Redes Genéticas	4
1.4. Enfoques Actuales	4
1.4.1. Sistemas de Ecuaciones Diferenciales Ordinarias con Valores Reales	4
1.4.2. Modelos Estáticos	5
1.4.3. Primera Aproximación a un Modelo Dinámico	6
1.5. Observaciones Finales	8
2. GENIE	9
2.1. Introducción	9
2.2. Construcción del Sistema de ODE's	10
2.2.1. Feedback Positivo de un Gen	13
2.2.2. Fosforilación-Defosforilación de una Proteína	15
2.3. Simulaciones	18
2.4. Marco Conceptual	22
2.4.1. Realismo	22
2.4.2. Flexibilidad y Completitud Funcional	26
2.4.3. Robustez de los Parámetros	27
2.4.4. Robustez de las Concentraciones Iniciales	29
2.5. Algoritmo GENIE	30
2.5.1. Diagrama de Flujo	30
2.5.2. Algoritmo	32
2.6. Observaciones Finales	37

3. Aplicación de GENIE a una Red Genética	39
3.1. Introducción	39
3.2. Modelado Basado en Binding Sites	41
3.3. Resultados Obtenidos de la Aplicación de GENIE	43
3.3.1. Iteración 1: Reducido	45
3.3.2. Iteración 2: Reducido sin Feedback	53
3.3.3. Iteración 3: Alternativa 1	55
3.3.4. Iteración 4: Alternativa 2.1	57
3.3.5. Iteración 5: Alternativa 2.2	60
3.3.6. Iteración 6: Alternativa 2.3	62
3.3.7. Iteración 7: Alternativa 2.4	64
3.3.8. Iteración 8: Enriquecido	66
3.4. Resumen de los Resultados Numéricos	84
3.4.1. Realismo	84
3.4.2. Flexibilidad	90
3.5. Predicciones y Validación Experimental	92
3.5.1. Binding: Experimentos ChIP	92
3.5.2. Expresión: Experimentos de Real-Time PCR	99
3.5.3. Materiales y Métodos	101
3.6. Observaciones Finales	101
4. Conclusiones	103
4.1. Introducción	103
4.2. GENIE Predice Interacciones	103
4.2.1. El Modelo Enriquecido Predice Exhaustivamente la Funcionalidad de la Red	105
4.3. PmrA/PmrB y PhoP/PhoQ Constituyen una Red Genética Robusta y Flexible	106
4.4. La Autoregulación Positiva Amplifica la Señal Transducida	107
4.5. La Transferencia de Información Es Soportada por la Fosfori- lación	107
4.6. PmrA/PmrB y PhoP/PhoQ Presentan una Funcionalidad Co- ordinada	107
4.7. La Agregación de los Módulos PmrA/PmrB y PhoP/PhoQ Genera una Red Genética	108
4.8. Activación coordinada de pbgP	109
4.9. GENIE Predice Nuevos Grupos de Genes	110
4.10. Generación Adaptativa de Hipótesis	111
4.11. Trabajos Futuros	112

Índice de figuras

1.1. Esquema genérico de funcionamiento de los sistemas de dos componentes PmrA/PmrB y PhoP/PhoQ	3
2.1. Esquema de funcionamiento biológico para la autoregulación positiva de la transcripción del gen phop	14
2.2. Arquitectura del módulo de autoregulación positiva	14
2.3. Ecuaciones diferenciales para el módulo de autoregulación positiva	14
2.4. Esquema biológico para las reacciones de fosforilación-defosforilación de PHOP	16
2.5. Arquitectura del módulo para las reacciones de fosforilación-defosforilación de PHOP mediadas por PHOQ.	16
2.6. Ecuaciones diferenciales para el módulo de fosforilación - defosforilación	17
2.7. Resultados de una corrida simulación realizada para una arquitectura para un cierto conjunto de parámetros.	19
2.8. Resultados esperados de una corrida simulación de una arquitectura para ciertas especies seleccionadas.	20
2.9. Diagrama de flujo para el algoritmo GENIE.	31
3.1. Regiones promotoras de los genes mgtA, mgtC y PmrD.	41
3.2. Interacciones entre los genes de los subsistemas PhoP/PhoQ y PmrA/PmrB derivadas de la información obtenida del estudio de los binding sites.	42
3.3. Modelo Reducido.	46
3.4. Modificaciones al esquema de la figura para el modelo Reducido que determinan la arquitectura del modelo Reducido sin Feedback.	53
3.5. Modificaciones al esquema de la figura para el modelo Reducido que determinan la arquitectura del modelo Alternativa 2.1.	58
3.6. Modificaciones al esquema de la figura para el modelo Reducido que determinan la arquitectura del modelo Alternativa 2.2	60
3.7. Modificaciones al esquema de la figura para el modelo Reducido que determinan la arquitectura del modelo Alternativa 2.3.	62

3.8. Modelo Enriquecido.	69
3.9. Robustez de los parámetros.	77
3.10. Robustez de los parámetros (continuación).	78
3.11. Robustez de los parámetros (continuación).	79
3.12. Proporción de cada tipo de parámetro que sólo puede tomar valores en un intervalo inferior al 25 % del intervalo biológicamente significativo, por constraint para el modelo Enriquecido.	80
3.13. Proporción de cada tipo de parámetro que puede tomar valores en el 100 % del intervalo biológicamente significativo, por constraint para el modelo Enriquecido.	81
3.14. Proporción de cada tipo de parámetro que sólo puede tomar valores en un intervalo inferior al 25 % del intervalo biológicamente significativo para la conjunción de los constraints en el modelo Enriquecido.	82
3.15. Proporciones para los modelos Reducido y Reducido sin Feedback.	86
3.16. Probabilidades para los modelos Reducido y Reducido sin Feedback.	86
3.17. Proporciones para los modelos Alternativa 2.	87
3.18. Probabilidades para los modelos Alternativa 2.	88
3.19. Proporciones para el modelo Enriquecido.	89
3.20. Probabilidades para el modelo Enriquecido.	89
3.21. Proporciones para la conjunción de los constraints.	91
3.22. Probabilidades para la conjunción de los constraints.	91
3.23. Curvas de expresión obtenidas para genes regulados por PhoP al modificar los valores de los parámetros de su ecuación diferencial asociada.	95
3.24. Binding de PhoP a la región promotora de distintos genes en función del tiempo.	96
3.25. Clustering jerárquico sobre los datos de los experimentos de CHIP.	97
3.26. Resultados de expresión para los genes <i>mgtA</i> , <i>pmrD</i> y <i>mgtC</i> medidos mediante Real-Time PCR.	99

Índice de cuadros

2.1. Explicación del significado de los parámetros utilizados en las ecuaciones diferenciales	12
2.2. Ejemplo de funcionalidad a reproducir por una arquitectura.	23
3.1. Resumen de las arquitecturas obtenidas para la red genética PhoP/PhoQ - PmrA/PmrB.	44
3.2. Patrones de entrada-salida para el modelo Reducido.	47
3.3. Términos de las ecuaciones que determinan las concentraciones de las especies.	48
3.4. Resultados para el modelo Reducido.	51
3.5. Resultados para la conjunción de los constraints 1 & 2 & 3 & 4 para el el modelo Reducido.	52
3.6. Patrones de entrada-salida para el modelo Reducido Sin Feedback.	53
3.7. Resultados para el modelo Reducido sin Feedback.	54
3.8. Resultados para la conjunción de los constraints 1 & 2 & 3 & 4 para el modelo Reducido sin Feedback.	54
3.9. Patrones de entrada-salida para el modelo Alternativa 1.	55
3.10. Resultados para el modelo Alternativa 1.	56
3.11. Resultados para la conjunción de los constraints 1 & 2 & 3 & 4 para el modelo Alternativa 1.	56
3.12. Resultados para el modelo Alternativa 2.1.	59
3.13. Resultados para la conjunción de los constraints 1 & 2 & 3 & 4 para el modelo Alternativa 2.1.	59
3.14. Resultados para el modelo Alternativa 2.2.	61
3.15. Resultados para la conjunción de los constraints 1 & 2 & 3 & 4 para el modelo Alternativa 2.2.	61
3.16. Resultados para el modelo Alternativa 2.3.	63
3.17. Resultados para la conjunción de los constraints 1 & 2 & 3 & 4 para el modelo Alternativa 2.3.	63
3.18. Resultados para el modelo Alternativa 2.4.	64
3.19. Resultados para la conjunción de los constraints 1 & 2 & 3 & 4 para el modelo Alternativa 2.4.	65
3.20. Patrones de entrada-salida para el modelo Enriquecido	70

3.21. Términos de las ecuaciones que determinan las concentraciones de las especies.	71
3.22. Resultados para el modelo Enriquecido.	74
3.23. Resultados para la conjunción de los constraints 1 & 2 & 3 & 4 para el modelo Enriquecido.	75
3.24. Proporción de cada parámetro particular que sólo puede tomar valores en un intervalo inferior al 25 % del intervalo biológicamente significativo.	81
3.25. Proporción de soluciones válidas al modificar los valores iniciales de las especies.	83
3.26. Resultados del los tests de realismo para todos los modelos.	85
3.27. Resultados para la conjunción de los constraints 1 & 2 & 3 & 4.	90
3.28. Ocupación relativa de phoP en los sitios de transcripción de phoP, mgtA y pmrD en condiciones de bajo magnesio y bajo hierro (constraint 2).	92
3.29. Correlación entre las curvas obtenidas en las simulaciones y las obtenidas en experimentos de inmunoprecipitación que determinan la ocupación relativa de phoP en los sitios de transcripción de phoP, mgtA y pmrD.	93
3.30. Curvas de expresión predichas por GENIE para distintos genes regulados por PhoP al modificar los parámetros de la solución.	94
3.31. Correlación entre las curvas predichas por GENIE y las curvas de binding obtenidas experimentalmente para distintos genes regulados por PhoP.	98
3.32. Correlación entre las curvas predichas y las curvas de expresión obtenidas experimentalmente.	100

Agradecimientos

El autor quiere realizar un agradecimiento a las siguientes personas, sin las cuales este trabajo no hubiera sido posible.

A Igor, cuya visión y conocimientos son un estímulo constante, por ser el guía en mis primeros pasos en la bioinformática.

A Emacs, Gaby y Juancito, compañeros de equipo inseparables, en compañía de los cuales espero seguir afrontando nuevos desafíos.

A Caro, por soportarme durante la realización de este trabajo, y por muchas cosas más.

Al Departamento de Computación, por su apoyo para la realización de este trabajo.

Al equipo del Groisman Lab de la Washington University in St. Louis, y especialmente a A. Kato y S. Dongwoo por la realización de los experimentos de biología molecular utilizados en este trabajo y a E. Groisman por la lectura y corrección de los modelos presentados.

A todos los miembros del Departamento de Computación, por hacer de mi carrera una enriquecedora y agradable experiencia.



Capítulo 1

Redes Genéticas

1.1. Introducción

Es posible afirmar que el gran desafío de la era postgenómica puede ser sintetizado respondiendo preguntas del tipo *cuándo, dónde, y por cuánto tiempo* [Bre00] un gen está prendido o apagado. En la actualidad existe un alto grado de incertidumbre en cuanto a los mecanismos precisos que determinan las respuestas a estas preguntas, siendo un objetivo principal de este trabajo el desarrollar un sistema de inferencia que permita formular y validar hipótesis para responderlas.

En la sección 1.2 se describen los problemas biológicos asociados al estudio de las redes genéticas. En la sección 1.3 se enuncian las distintas problemáticas referentes al modelado computacional de las redes genéticas. En la sección 1.4 se realiza un análisis de los enfoques de mayor relevancia para el presente trabajo, elaborándose conclusiones acerca de las fortalezas y debilidades de los mismos.

1.2. Problema Biológico

1.2.1. Sistemas de Dos Componentes

Una pregunta fundamental en biología es cómo un organismo integra múltiples señales para generar una respuesta celular apropiada. Para cada conjunto de señales, existe un cierto número de genes (no siempre conocido en su totalidad) que codifican las distintas especies proteicas encargadas de capturar el estímulo (input), realizar la transducción de señales necesaria para activar la maquinaria celular de respuesta, y finalmente generar la respuesta adecuada (output), generalmente resumida en una variación en la expresión de nuevos genes. Estos conjuntos de genes determinan verdaderos circuitos biológicos de control celular.

En los procariotas, una importante cantidad de funciones celulares es

controlada por sistemas regulatorios de dos componentes. Circuitos dedicados transducen e interpretan señales específicas tales como PH, temperatura, osmolaridad, luz, nutrientes, iones, feromonas, y toxinas para regular un amplio rango de procesos que incluyen movilidad, virulencia, metabolismo, ciclo celular, switches de desarrollo, resistencia a antibióticos y respuesta a stress [HS95]. Dos modelos biológicos ampliamente estudiados son los sistemas de dos componentes PmrA/PmrB y PhoP/PhoQ, tanto en su funcionamiento individual como en su interacción [KLG03]. El caso de la acción coordinada de estos sistemas reviste especial interés dado que presenta las características de una verdadera *red genética*, conservando una relativa simpleza estructural.

1.2.2. Los Sistemas de Dos Componentes PhoP/PhoQ y PmrA/PmrB

Los sistemas de dos componentes constituyen el mecanismo principal de transducción de señales que permiten a las bacterias modificar su comportamiento celular en respuesta a estímulos del ambiente. Estos sistemas consisten en una proteína sensora con actividad histidin-kinasa que responde a señales específicas fosforilando a su proteína reguladora de respuesta asociada. La asociación con un ligando específico (estímulo) modifica tres actividades enzimáticas de la proteína sensora: la capacidad de autofosforilarse en presencia de ATP (autokinasa), la transferencia de este grupo fosforil a su regulador asociado (kinasa) y la defosforilación de éste último (fosfatasa). En su mayoría, los reguladores de respuesta son factores de transcripción capaces de asociarse al DNA de genes específicos, siendo la fuerza de esta asociación modificada por la fosforilación. De esta manera, la fosforilación o defosforilación de estas proteínas puede modificar la expresión de ciertos genes de la bacteria, y una señal específica del medio se transduce en una señal de respuesta de la bacteria en la forma de una modificación en su perfil de expresión.

El sistema de dos componentes PmrA/PmrB (regulador de respuesta/sensor, respectivamente) presente en *Salmonella enterica serovar Typhimurium*, organismo tomado como modelo para la implementación del método GENIE, es necesario para la resistencia inducible al antibiótico polimyxina B, la resistencia a la muerte mediada por Fe³⁺, el crecimiento en suelo, la virulencia en ratones, y la infección de macrófagos de pollo. Responde independientemente a dos señales, alto nivel de Fe³⁺ extracelular, sentido por PmrB, y bajo nivel de Mg²⁺, sentido por la proteína PhoQ.

El sistema PhoP/PhoQ (regulador de respuesta/sensor, respectivamente), presente en el mismo organismo, constituye un regulador maestro (regulando aproximadamente la transcripción del 2% del genoma de la bacteria) que gobierna la adaptación a medios de bajo Mg²⁺ y la virulencia en ratones así como también otras funciones biológicas. Su funcionamiento como sistema de dos componentes es en principio análogo al anterior, siendo el

estímulo activador el bajo nivel de Mg^{2+} extracelular. Entre los genes activados por PhoP se encuentra el gen *pmrD*, el cual resulta de especial interés porque presenta un sitio de binding para PmrA, y su producto, la proteína PmrD, puede asociarse a la proteína PmrA. Ambos sistemas de dos componentes muestran una coordinación de sus funciones in vivo, aunque los mecanismos exactos de interacción son aún desconocidos. Un esquema de su funcionamiento puede apreciarse en la Figura 1.1.

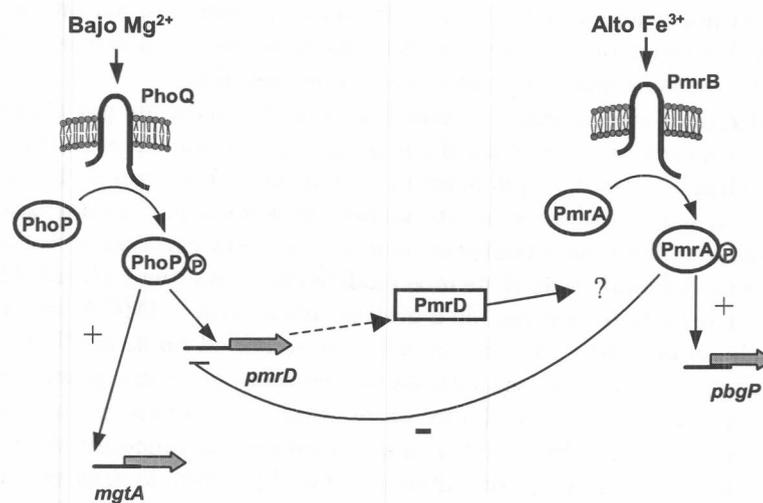


Figura 1.1: Esquema genérico de funcionamiento de los sistemas de dos componentes PmrA/PmrB y PhoP/PhoQ

1.3. Problema Computacional

1.3.1. Modelado de Redes Genéticas

Actualmente existe una gran cantidad de formas de modelar redes genéticas que interpretan señales y representan relaciones dinámicas entre genes[vSWEB02]. En principio, se intenta definir a la red como un grafo en el cual las distintas especies moleculares de la red son representadas por los nodos, y las interacciones entre estas especies son representadas por los ejes. Los valores asignados a los nodos (es decir, el valor de la concentración de la especie representada) y ejes (es decir, el valor la interacción entre dos especies) pueden definirse utilizando dominios booleanos, discretos o continuos; las interacciones entre los genes pueden ser consideradas estáticas o dinámicas, y el modelo en su conjunto puede ser estocástico o determinístico.

La calidad de la red obtenida no está determinada solamente por el modelo elegido, sino también por el diseño de un método de inferencia (estrategia de aprendizaje) adecuado que permita estimar los parámetros de la red. Una de las mayores dificultades en este sentido es causada por la dimensión del problema; las distintas estrategias para superar esta cuestión se basan en la reducción del número de elementos modelados (clustering y thresholding), el incremento de la cantidad de muestras (microarray)[MMCW00] o la simplificación de la complejidad del modelo (conectividad limitada)[WHA].

El desarrollo de modelos computacionales de redes genéticas posee una gran importancia, dado que posibilita, entre otras aplicaciones, desarrollar modelos dinámicos de sistemas vivientes que, mediante la aplicación de distintas técnicas de data mining, permiten acelerar el proceso de *drug discovery*, facilitando el descubrimiento y caracterización de blancos terapéuticos potenciales.

1.4. Enfoques Actuales para la Modelización Computacional

1.4.1. Sistemas de Ecuaciones Diferenciales Ordinarias con Valores Reales

Un caso particularmente interesante de modelado de redes genéticas es aquel en el que se utilizan valores continuos para determinar los niveles de expresión de los compuestos y de las relaciones entre los mismos. Esto permite capturar aspectos biológicos observados experimentalmente, que son sacrificados en pro de la simplicidad en los modelos booleanos como el propuesto en [FL98].

Una posible aproximación a estos modelos continuos es la basada en sistemas de ecuaciones diferenciales ordinarias (ODE's, del inglés Ordinary Differential Equation) no lineales, en la cual se proponen sistemas de ecuaciones

que permiten calcular la variación en las concentraciones de las especies, RNA o proteína (valor de un nodo), en función del tiempo a partir de los valores de los nodos conectados con él. De esta manera, pueden realizarse simulaciones temporales del sistema, permitiéndose observar aspectos dinámicos del mismo, esenciales en los sistemas de control. Mediante estas simulaciones, es posible extraer propiedades emergentes a nivel sistémico, tales como la robustez de la red.

La bibliografía existente sobre este tema puede dividirse según el uso de modelos “estáticos” o “dinámicos” de ODE’s. En los modelos de ODE’s estáticos, como el propuesto en [BG03a], se propone que el sistema biológico alcanza un estado de equilibrio (es decir, en el cual el diferencial de las concentraciones de los productos de sus genes en función del tiempo es igual a cero), en el cual se determinan las ecuaciones específicas del modelo, determinándose consiguientemente la red asociada. En los modelos dinámicos (como el propuesto en [vDMMO00a]) no se considera necesariamente este equilibrio, con lo que pueden observarse variaciones en los valores de expresión de los genes (valores de los nodos) a lo largo del tiempo. Esta última característica es la que permite realizar simulaciones temporales del comportamiento del sistema, siendo este hecho de especial interés al estudiar sistemas biológicos de regulación tales como PmrA/PmrB-PhoP/PhoQ, en los que puede observarse experimentalmente una respuesta dinámica a variaciones en los estímulos recibidos (variaciones en los valores de los nodos de entrada).

La noción de que las complejas redes de interacción molecular dentro de las células deberían ser robustas a variaciones de los parámetros que las definen aparece en distintos contextos dentro de la biología celular. En este sentido, suele sugerirse que esta robustez debería ser intrínseca a la arquitectura de la red, es decir que esta debería preservar sus características funcionales al producirse perturbaciones en uno o más de sus parámetros.

En las secciones siguientes se analizan distintos enfoques que permiten modelar redes genéticas en base a sistemas de ecuaciones diferenciales. En el primero, se utiliza un sistema de ecuaciones estático, en el sentido de que sólo se trabaja sobre aspectos matemáticos de las mismas (sin modificar sus parámetros) no realizándose pruebas que permitan determinar el comportamiento del sistema obtenido. En el segundo, se plantea un sistema dinámico que permite realizar simulaciones sistemáticas para determinar no sólo el comportamiento del sistema bajo distintas condiciones, sino también los valores de los parámetros de las mismas.

1.4.2. Modelos Estáticos

En modelos basados en ecuaciones diferenciales como el propuesto en [BG03b] para el sistema de dos componentes EnvZ/OmpR, se sugiere que la robustez de la red se evidencia en que el output de la misma no se ve

afectado significativamente al variar los valores de las concentraciones de las especies de entrada. En este contexto, un conjunto de nodos es distinguido como “nodos de entrada” en el sentido de que sus miembros representan la interacción entre el entorno y la red genética (funcionando como receptores de señales); otro conjunto es distinguido como “nodos de salida” significando que el valor de expresión de sus miembros representa la respuesta a las señales (funcionando como el efecto de la transducción de señales). De esta manera, se sugiere una *robustez con respecto a los valores iniciales de los nodos de entrada del sistema*. Este enfoque, sin embargo, adolece de los siguientes defectos:

- Los parámetros de las ecuaciones diferenciales que definen este tipo de modelos no son generalmente conocidos, por lo que deben ser estimados en forma relativamente arbitraria.
- Una vez obtenido el modelo, resulta complejo, cuando no imposible, validar estos valores experimentalmente.
- El modelo es estático, dado que se asume que el sistema se encuentra en un estado estable en el cual no hay variaciones temporales en las concentraciones, e.g. si bien EnvZ puede estar tanto fosforilado como defosforilado, se asume que la suma de las concentraciones de ambas especies permanece constante, no existiendo creación (síntesis) ni destrucción (decaimiento) de EnvZ. Esto impide detectar aspectos dinámicos del sistema, inherentes a todo proceso de procesamiento de señales, y entorpece asimismo la tarea de detectar una resistencia del sistema en el sentido de que su comportamiento no se vea afectado por variaciones en el valor de algunos de los nodos de la red.

1.4.3. Primera Aproximación a un Modelo Dinámico

En [vDMMO00b] se propone un sistema de ODE's no lineales, las cuales caracterizan el cambio en el tiempo de las concentraciones de las especies moleculares (valores de los nodos) presentes en la red, ya sean proteínas o RNA's. En base a esto, se realizan simulaciones temporales para medir el comportamiento del sistema. Un concepto interesante de este modelo es que los valores de los parámetros no son estimados a priori, sino que mediante distintos métodos (búsqueda al azar, por ejemplo) se generan conjuntos de parámetros que luego son probados verificando que el modelo reproduzca ciertos patrones macroscópicamente observables del modelo biológico estudiado.

En este enfoque, se proponen tres medidas de robustez:

- Con respecto a la solución obtenida, es decir con respecto a las *características funcionales del sistema*. Se propone que una cierta arquitectura de red es robusta cuando, al generar conjuntos de parámetros

aleatoriamente, presenta una proporción de soluciones válidas¹ que se desvía de lo esperable en función del azar.

- Con respecto a los valores de los parámetros, es decir respecto a los *valores de los ejes*. Se propone que un red es robusta si, una vez obtenida una solución válida, esta admite un amplio rango de variación para sus parámetros en forma individual manteniendo la (única) funcionalidad esperada. Se sugiere que esto podría simular la resistencia a pequeñas mutaciones en los genes.
- Con respecto a los valores de concentraciones iniciales, es decir los *valores iniciales de los nodos*. Se propone que una red es robusta si mantiene su patrón de actividad al variar las concentraciones iniciales de las especies, lo cual podría representar la resistencia de la red al “ruido” del entorno. No se presenta un desarrollo profundo de este concepto debido a las características puntuales del modelo en cuestión.

Si bien estas medidas de robustez son acertadas en el modelo puntual del artículo, resultan insuficientes al aplicarlas a un modelo de transducción de señales como EnvZ/OmpR, modelado en [BG03a], o PmrA/PmrB-PhoP/PhoQ, estudiado en [KLG03] y modelado en el presente trabajo. De acuerdo a la distinción de nodos de entrada y nodos de salida realizada anteriormente, este tipo de sistemas presenta distintos patrones de valores en sus nodos de salida en función del estímulo presentado, es decir en función de los patrones de valores presentes en los nodos de entrada.

En un trabajo posterior [MvDMO02a], se plantea una medida para determinar la *flexibilidad* de una red genética. En el modelo estudiado, coincidentemente con la red PmrA/PmrB-PhoP/PhoQ, resulta de interés determinar el poder de las soluciones obtenidas para reproducir distintos patrones de salida dependiendo del patrón inicial de concentraciones presentado (es decir, dependiendo del patrón de valores presente en los nodos iniciales). Se sugiere que esta flexibilidad es alcanzable evolutivamente gracias a la robustez intrínseca de la red, es decir que ambas propiedades sistémicas se encuentran evolutivamente acopladas. Sin embargo, no se orienta la búsqueda de soluciones a obtener modelos flexibles, sino que esto se reporta más bien como una propiedad obtenida colateralmente. Asimismo, no se asegura que los distintos patrones a reproducir sean exhaustivos ni mutuamente excluyentes.

¹A lo largo del texto denominaremos *solución válida* a un conjunto de parámetros que hace que la red pueda reproducir un cierto comportamiento deseado, como por ejemplo una serie de valores de concentraciones en los nodos de salida al inicializarse los nodos de entrada en un valor específico.

1.5. Observaciones Finales

En los enfoques expuestos se presentan distintas medidas para determinar la robustez y flexibilidad de una red genética en base a modelos computacionales. Sin embargo, estos enfoques son parciales y no presentan una medida integral acerca de la “calidad” de la red propuesta para representar un modelo biológico. En el presente trabajo se propone una medida de calidad, basada en los criterios que se siguen a continuación.

- **Realismo:** la red debe ser capaz de reproducir el comportamiento observable experimentalmente, es decir el funcionamiento biológico del sistema. Una red realista en este sentido debe ser capaz de reproducir este funcionamiento en forma relativamente independiente de los parámetros escogidos (por ejemplo, al generar aleatoriamente conjuntos de parámetros, la red debe reproducir el comportamiento biológico en un proporción que se desvíe de lo esperable en función del azar).
- **Robustez respecto de**
 - **parámetros (ejes):** la red debe soportar variaciones de distinto rango en parámetros particulares, sin que se vea afectado su realismo. Esto representa la propiedad biológica de resistencia a pequeñas mutaciones en los genes de la red.
 - **concentraciones (nodos):** la red debe soportar variaciones en los valores de los nodos de entrada, sin que se vea afectado su realismo. Esto representa la resistencia de la red al ruido presente en los sistemas moleculares.
- **Flexibilidad:** la red debe ser capaz de reproducir simultáneamente los distintos patrones de comportamiento observados (o propuestos) en el sistema biológico en estudio. Este criterio se encuentra íntimamente relacionado con la robustez, en el sentido de que es plausible proponer que la flexibilidad comúnmente observada en las redes genéticas es alcanzable en sistemas robustos que soporten la acumulación de mutaciones que producirá, a lo largo de la evolución, el surgimiento de nuevos patrones funcionales.

Finalmente, ninguno de los trabajos analizados propone un método que permita la formalización automática de un problema biológico mediante la generación de arquitecturas de red que satisfagan las reglas biológicas que lo definen. Tampoco se propone un mecanismo que permita reformular estas reglas, es decir generar nuevas hipótesis sobre el sistema, en función de los resultados obtenidos.

Capítulo 2

GENIE: Gene Expression Networks Iterative Explorer

2.1. Introducción

En este capítulo se propone un modelo para la inferencia de redes genéticas, GENIE¹, basado en un análisis secuencial e iterativo de las propiedades de las arquitecturas enunciadas en la sección anterior, realismo, flexibilidad, y robustez de nodos y ejes. En cada una de las secciones se analiza en particular cada uno de los elementos que constituyen la metodología de trabajo GENIE, siendo estos: identificación de reglas en base a binding sites y bibliografía, identificación de arquitecturas consistentes con estas reglas, evaluación de realismo, flexibilidad y robustez de las arquitecturas obtenidas, reformulación de las mismas, generando hipótesis adaptativamente, realización de predicciones y contrastación experimental biológica.

En la sección 2.2 se presentan las técnicas utilizadas para traducir el problema estudiado del dominio biológico al dominio computacional, traduciendo reglas extraídas de la literatura y evidencia de binding sites en arquitecturas de redes genéticas implementadas mediante sistemas de ecuaciones diferenciales, y testeadas en su comportamiento dinámico mediante simulaciones temporales. En la sección 2.3 se describen las simulaciones realizadas utilizando el sistema de ecuaciones diferenciales obtenido. En la sección 2.4 se describen los elementos conceptuales utilizados para formular la conexión entre las técnicas de modelización desarrolladas y el algoritmo presentado. En base a estos elementos, en la sección 2.5 se formula el algoritmo GENIE, mostrándose un diagrama de flujo para el mismo y una implementación en pseudocódigo. Este algoritmo permite formalizar el pro-

¹Voz inglesa para el español *genio*, derivada del latín *genius*. En la mitología, deidad de la Naturaleza, tenida por principio engendrador de todas las cosas. En una de sus acepciones [Esp01] “capacidad mental extraordinaria para crear o inventar cosas nuevas y admirables”.

blema biológico en estudio, refinando la modelización en base a propiedades sistémicas emergentes y realizando distintas predicciones durante el aprendizaje del modelo

2.2. Construcción del Sistema de ODE's² No Lineales para las Arquitecturas

En base a las reglas para el funcionamiento de la red genética estudiada recogidas de la literatura y la evidencia de los sitios de binding de los genes involucrados, es posible proponer una o varias arquitecturas para el sistema. Estas arquitecturas modelan las interacciones entre las especies, y pueden visualizarse como grafos en los cuales los nodos representan las especies involucradas (proteínas y RNA), y los ejes representan las interacciones entre las mismas. Para obtenerlas, se utilizan librerías de módulos funcionales como los descritos en [LR02], [MSOI⁺02] y [Kae03], definiéndose a la red como una agregación de submódulos, como por ejemplo autoregulación positiva o negativa de un gen, regulación de un gen en base a los productos de un grupo de genes, fosforilación y defosforilación de una proteína mediada por una cierta enzima, etc.

Particularmente, las propiedades que caracterizan la implementación seguida por GENIE son las siguientes:

- Los submódulos obtenidos son traducidos utilizando librerías de ecuaciones diferenciales como las enumeradas en [MMOD02], [vDMMO00a] y [Kae03]. De esta manera, el valor de los ejes de la arquitectura queda determinado por las ecuaciones propuestas como una función de ciertos parámetros bioquímicos. Estos parámetros no son en general conocidos, y por lo tanto son estimados computacionalmente. Si bien existen enfoques para la inferencia de arquitecturas que utilizan evidencia experimental para fijar los parámetros [BG03a], la cantidad de los mismos en cualquier red de modesta dimensión hace imposible determinarlos experimentalmente en su totalidad; como se verá posteriormente, esta fijación resulta inclusive perjudicial dado que impide extraer información acerca del papel de su variación dentro del funcionamiento de la red.
- Las ecuaciones obtenidas para los diferenciales en función del tiempo de las especies permiten calcular, en un tiempo dado, la concentración de una especie (valor de un nodo) como una función de los valores de las concentraciones de las especies (nodos) conectadas con ella. Esto permite la realización de simulaciones temporales sobre el sistema de ODE's obtenido, posibilitando capturar el comportamiento dinámico del sistema.

²Ordinary Differential Equations

- Tanto los parámetros utilizados como las concentraciones son adimensionales, por lo que no resulta necesario realizar normalizaciones de unidades en las ecuaciones. Las concentraciones de las especies toman valores reales entre 0 (nulo) y 1 (máximo). Las conversiones algebraicas utilizadas para lograr la adimensionalización se describen en [vDMMO00a].

A continuación se presentan dos ejemplos del mecanismo de construcción, enunciados en forma general. Si bien esta esquematización es válida en general, con el fin de ejemplificar el modelado del problema particularmente estudiado en este trabajo se mostrara una instanciación de los módulos genéricos con especies del sistema PhoP/PhoQ. Para denominar a las especies en los diagramas de arquitectura, utilizaremos letras mayúsculas para los nombres de las proteínas y minúsculas para los RNA. El DNA no se modela explícitamente, y por lo tanto la interacción de, por ejemplo, una proteína activadora sobre un cierto gen se muestra en la arquitectura mediante un eje entre la especie proteica y la especie RNA, simbolizando que la cantidad de RNA puede verse como una función de la cantidad de activador presente.

Las ecuaciones diferenciales utilizadas para modelar las reacciones contienen parámetros de distinta significación biológica. Con el fin de aumentar la declaratividad de las ecuaciones, se utilizó la siguiente convención de nombres: un prefijo (e.g. H_{-} , nu_{-}) identifica la función biológica del parámetro (vida media, coeficiente de cooperatividad de Hill), y luego se siguen los nombres de las especies involucradas. Por ejemplo nu_PHOQ_PHOPP es el parámetro que determina el coeficiente de asociatividad de Hill para las reacciones entre PHOQ y PHOPP (PHOP fosforilado). Esta nomenclatura se resume en el Cuadro 2.1, en el cual se indican los rangos de valores considerados válidos en la modelización.

Prefijo	Significado	Rango de valores utilizado.
K	Coefficiente de activación medio-maximal. Equivale a la concentración en la cual la reacción presenta la mitad de la actividad máxima.	0,001 - 1
H	Vida media (inversa de la tasa de degradación).	1 - 100
nu	Coefficiente de Hill, valores cercanos a 1.0 producen curvas casi lineales, incrementándose la forma de s (sigmoidea) al aumentar este parámetro.	1 - 10
alpha	Coefficiente de saturación para un enhancer.	0.1 - 10
r	Tasa de transformación, por ejemplo para la reacción de transformación de PHOQ a PHOQ-ACT	0,001 - 10
P	Tasa de fosforilación	0,001 - 10
T_0	Constante de tiempo característica, derivada del reemplazo utilizado para llevar el sistema de ecuaciones a una forma no dimensional. Relaciona al tiempo dimensional (t) con el tiempo adimensional (τ): $t = T_0\tau$.	

Cuadro 2.1: Explicación del significado de los parámetros utilizados en las ecuaciones diferenciales

2.2.1. Ejemplo de Modelo Dinámico: Feedback Positivo de un Gen

Uno de los módulos más sencillos utilizados en GENIE es el de la autoregulación positiva o feedback positivo de un gen. Las reglas que lo definen son:

- “La proteína X se asocia al DNA del gen X, aumentando la transcripción del RNA de X”
- “El RNA de X es traducido a la proteína X”
- “El RNA X sufre decaimiento espontáneo”³
- “La proteína X sufre decaimiento espontáneo”

Estas reglas, que definen el módulo de autoregulación positiva, resultan válidas para el gen *phoP* en el sistema PhoP/PhoQ, como se ve a continuación:

- “La proteína PhoP se asocia al DNA del gen *phoP*, aumentando la transcripción del RNA de PhoP”
- “El RNA de *phoP* es traducido a la proteína PhoP”
- “El RNA de *phoP* sufre decaimiento espontáneo”
- “La proteína PhoP sufre decaimiento espontáneo”

Estas reglas se basan en un mecanismo biológico como el que se esquematiza en la Fig.2.1, y determinan el módulo de “autoregulación positiva de un gen”, presente en la librería de módulos, cuya arquitectura se muestra en la Fig.2.2. En esta la existencia de una relación entre dos especies se determina en base a la existencia de una en el mecanismo biológico, y la calidad de esta relación es definida en función del tipo de interacción propuesto, utilizando la ecuación diferencial que la describe, escogida de la librería de ecuaciones. Luego estas ecuaciones se transforman en términos a adicionar o sustraer dentro de las ecuaciones diferenciales para cada una de las especies involucradas, como se ve en las Ecuaciones 2.3.

³Este decaimiento, que denominaremos decaimiento de primer orden, sucede espontáneamente in vivo para todas las especies

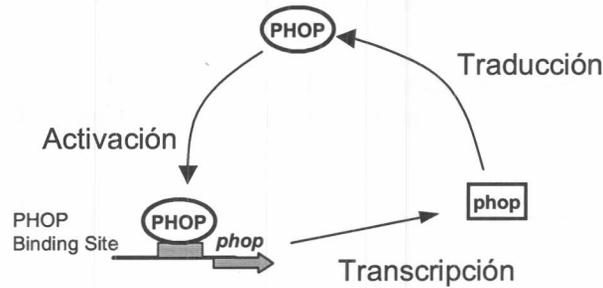


Figura 2.1: Esquema de funcionamiento biológico para la autoregulación positiva de la transcripción del gen *phop*
 Con el fin de mantener la simplicidad del ejemplo, se muestra la opción más sencilla para la autoregulación positiva de PHOP.

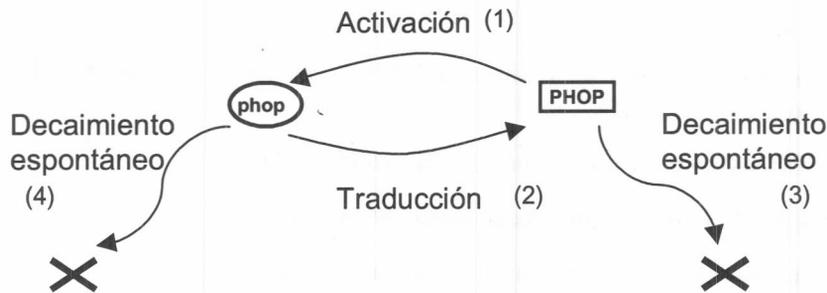


Figura 2.2: Arquitectura del módulo de autoregulación positiva
 Los números en los ejes referencian a los términos de las ecuaciones diferenciales derivadas.

$$\frac{d[phop]}{dt} = \frac{T_0}{H_{phop}} \frac{[PHOP]^{\nu_{PHOP-phop}}}{K_{PHOP-phop}^{\nu_{PHOP-phop}} + [PHOP]^{\nu_{PHOP-phop}}} - \frac{T_0[phop]}{H_{phop}} \quad (1)$$

$$\frac{d[PHOP]}{dt} = \frac{T_0}{H_{PHOP}} [phop]^{(2)} - \frac{T_0}{H_{PHOP}} [PHOP]^{(3)}$$

Figura 2.3: Ecuaciones diferenciales para el módulo de autoregulación positiva
 La numeración entre paréntesis utilizada para cada término se corresponde con la de la reacción del diagrama de arquitectura de la cual se deriva.

2.2.2. Fosforilación-Defosforilación de una Proteína

Otro módulo presente en las librerías utilizadas es el de fosforilación-defosforilación de una proteína. Este se enunciará para una cierta proteína X y una enzima Y , y luego se instanciará, sin pérdida de generalidad, para especies presentes en el sistema PhoP/PhoQ, resultando válida esta instanciación para todo sistema de dos componentes. Siguiendo el esquema de las secciones anteriores, se modelan las reacciones de fosforilación y defosforilación de una proteína X , tanto espontáneas como generadas por la actividad enzimática de la enzima Y . Las reglas a considerar serían:

- “ X es fosforilado a $X-P$ por la acción kinasa de Y ”
- “ $X-P$ es defosforilado por la acción fosfatasa de Y ”
- “ $X-P$ presenta un cierto nivel de defosforilación espontánea”
- “La proteína X sufre decaimiento espontáneo”
- “La proteína $X-P$ sufre decaimiento espontáneo”

La instanciación para el sistema PhoP/PhoQ resulta:

- “PhoP es fosforilado a PhoP-P por la acción kinasa de PhoQ”
- “PhoP-P es defosforilado por la acción fosfatasa de PhoQ”
- “PhoP-P presenta un cierto nivel de defosforilación espontánea”
- “La proteína PhoP sufre decaimiento espontáneo”
- “La proteína PhoP-P sufre decaimiento espontáneo”

En la Figura 2.4 se muestra el esquema de comportamiento biológico para el módulo, utilizando el ejemplo para el sistema PhoP/PhoQ. En la Figura 2.5 se muestra la arquitectura definida para la modelización, y en la Figura 2.6 se muestra el sistema de ecuaciones diferenciales obtenido.

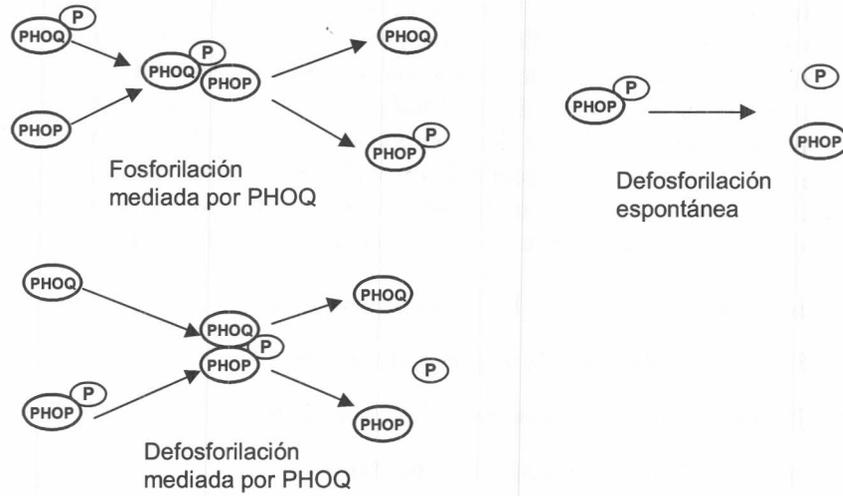


Figura 2.4: Esquema biológico para las reacciones de fosforilación-defosforilación de PHOP

Para preservar la simplicidad del ejemplo, se modela el esquema más sencillo de fosforilación de PHOP. No se agregan opciones exploradas por el método, como la que distingue dos formas de PHOQ, una normal con alto Mg^{2+} y una activada con bajo Mg^{2+} . Para esta opción, se explorará la combinatoria entre estos esquemas de fosforilación y las enzimas potenciales, PHOQ y PHOQ-ACT (ivado).

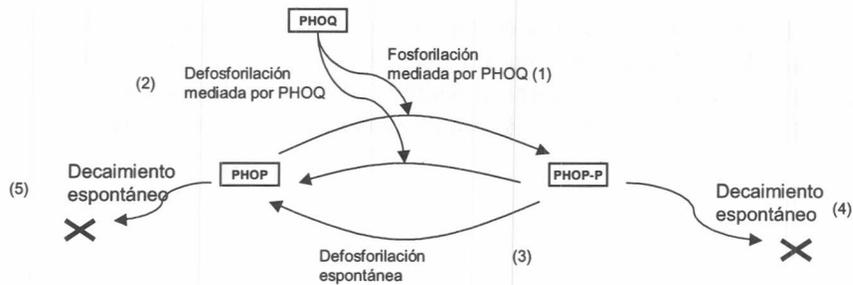


Figura 2.5: Arquitectura del módulo para las reacciones de fosforilación-defosforilación de PHOP mediadas por PHOQ.

Los números en los ejes referencian a los términos de las ecuaciones diferenciales derivadas.

$$\begin{aligned}
\frac{d[PHOP]}{dt} &= T_0(P_{PHOQ}[PHOPP]) \frac{[PHOQ]^{\nu_{PHOQ-PHOPP}}}{K_{PHOQ-PHOPP}^{\nu_{PHOQ-PHOPP}} + [PHOQ]^{\nu_{PHOQ-PHOPP}}} \quad (2) + [PHOPP]r_{PHOPP}^{(3)} \\
&\quad - P_{PHOQ}[PHOP] \frac{[PHOQ]^{\nu_{PHOQ-PHOP}}}{K_{PHOQ-PHOP}^{\nu_{PHOQ-PHOP}} + [PHOQ]^{\nu_{PHOQ-PHOP}}} \quad (1) - \frac{[PHOP]}{H_{PHOP}} \quad (5) \\
\frac{d[PHOPP]}{dt} &= T_0(P_{PHOQ}[PHOP]) \frac{[PHOQ]^{\nu_{PHOQ-PHOP}}}{K_{PHOQ-PHOP}^{\nu_{PHOQ-PHOP}} + [PHOQ]^{\nu_{PHOQ-PHOP}}} \quad (1) \\
&\quad - P_{PHOQ}[PHOPP] \frac{[PHOPP]^{\nu_{PHOQ-PHOPP}}}{K_{PHOQ-PHOPP}^{\nu_{PHOQ-PHOPP}} + [PHOPP]^{\nu_{PHOQ-PHOPP}}} \quad (2) - [PHOPP]r_{PHOPP}^{(3)} - \frac{[PHOPP]}{H_{PHOPP}} \quad (4)
\end{aligned}$$

Figura 2.6: Ecuaciones diferenciales para el módulo de fosforilación - defosforilación

La numeración entre paréntesis utilizada para cada término se corresponde con la de la reacción del diagrama de arquitectura de la cual se deriva.

2.3. Simulaciones Realizadas sobre las Arquitecturas

Los sistemas de ecuaciones diferenciales generados permiten calcular la concentración de cada una de las especies de la red para un instante dado del tiempo en función de las concentraciones de las mismas en el instante anterior. Para realizar la simulación, se asigna un valor inicial a todas las especies para el tiempo T_0 , y luego el motor de simulación calcula, en base a la integración de las ecuaciones, el valor de cada una de estas especies para los tiempos T_1, T_2, \dots, T_n . El ejemplo que se utiliza a continuación para describir una simulación y los criterios de éxito de la misma corresponde al utilizado para la red genética PhoP/PhoQ y PmrA/PmrB, aunque resulta válido para cualquier simulación realizada para una red genética genérica.

Utilizando un conjunto de valores de concentraciones iniciales definido (por ejemplo, Fe³⁺ alto, con valor 1.0, Mg²⁺ bajo, con valor 0.0, PMRA y PMRB 0.2, etc.), es posible calcular las curvas de concentración para todas las especies como se muestra en la Fig.2.7. Esto puede realizarse, para una misma arquitectura, con distintos conjuntos de parámetros de las ecuaciones diferenciales. Este hecho resulta importante dado que los valores de estos parámetros no son, en la mayoría de los casos, conocidos, y al independizarse de los mismos GENIE permite realizar predicciones sobre la red genética estudiada sin necesidad de fijarlos arbitrariamente. La simulación es considerada "exitosa" (es decir, se ha obtenido una "solución válida" para la red propuesta) si las curvas obtenidas para ciertas especies distinguidas de RNA, en nuestro ejemplo **mgta**, **pbgp** y **pmrd**, son similares a las que se obtendrían al generarse un cierto patrón esperado, como por ejemplo "mgta, pbgp y pmrd con concentración mayor a 0.8 a partir de un cierto instante en la simulación", determinando la activación o no de los genes correspondientes. Esto puede apreciarse en al Fig.2.8.

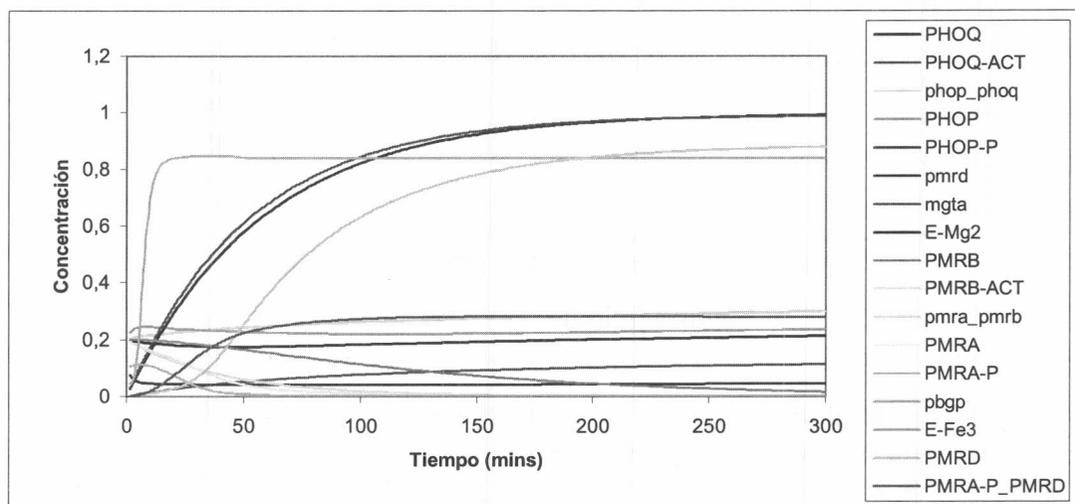


Figura 2.7: Resultados de una corrida simulación realizada para una arquitectura para un cierto conjunto de parámetros.

En la gráfica se muestra la variación de las concentraciones de las especies de la red a lo largo del tiempo de la simulación.

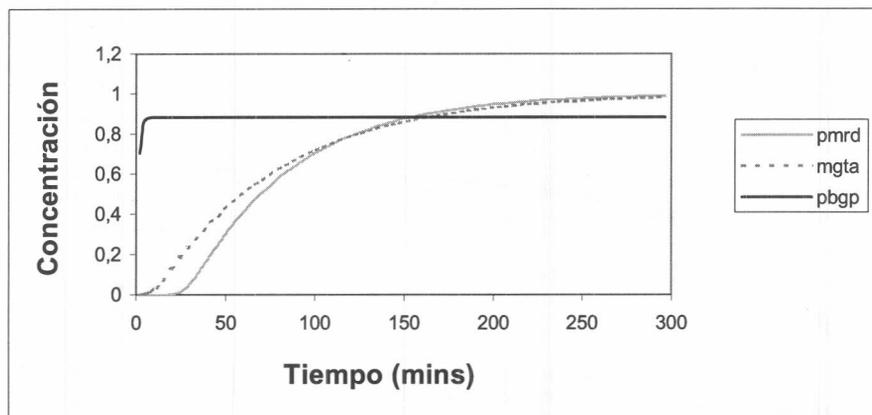


Figura 2.8: Resultados esperados de una corrida simulación de una arquitectura para ciertas especies seleccionadas.

Con el fin de determinar si una arquitectura, con un cierto conjunto de valores de sus parámetros, reproduce el comportamiento del sistema biológico, se proponen patrones de valores de la concentraciones a lo largo del tiempo que la simulación debe reproducir para ser considerada exitosa. En el gráfico se ve el patrón de expresión esperable para el sistema en condiciones de alto Fe^{3+} y bajo Mg^{2+} , en el cual se ve un aumento en la concentración de los RNA de mgta y pbgp, y una disminución del nivel del RNA pmrd. Cuanto más similares sean las curvas obtenidas por una cierta simulación a las del gráfico, mayor será el éxito de la arquitectura con el conjunto de parámetros escogido.

Para realizar las simulaciones, se realizaron pruebas con distintos integradores: Cash Karp, SEBE (“Semi Explicit Backward Euler”), SEAPC (“Semi-Explicit Adams Predictor-Corrector”) y Adams-Bashforth-Moulton. Los resultados finales presentados corresponden a la utilización de Cash Karp (para un análisis comparativo de estos métodos para un problema biológico como el estudiado, c.f.[MMOD02]).

Para determinar la habilidad de una cierta arquitectura, con su correspondiente conjunto de parámetros, para reproducir el comportamiento del sistema en el organismo vivo, se utilizó una función de score basada en los valores de las concentraciones de ciertas especies distinguidas⁴ Un score cercano a 0 indica una gran similitud con el constraint impuesto, y un valor cercano a 1 indica una gran diferencia. El score se calcula como la combinación de distintas funciones de umbral. Básicamente, se utilizan funciones sigmoideas para testear la expresión de las especies seleccionadas, por ejemplo considerando que una especie se encuentra activada si supera el 10% de su valor maximal de expresión, y apagada en caso contrario. Cada función de umbral utilizada devuelve un valor escalar con el valor medio-maximal en un umbral de acuerdo a las fórmulas:

$$T_{off} = \alpha_{max} \frac{(x_i/x_t)^3}{1 + (x_i/x_t)^3}$$

$$T_{on} = \alpha_{max} \left(1 - \frac{(x_i/x_t)^3}{1 + (x_i/x_t)^3}\right)$$

donde x_t representa el umbral para la especie x , α_{max} es el peor valor posible (0.5 en este ejemplo), y el subíndice i representa la especie i . La razón para utilizar sigmoideas es que estas funciones responden linealmente cerca del umbral pero proveen tanto penalizaciones como premios decrecientes al alejarse del umbral. Este hecho resulta coherente con la intuición de que si un gen particular se encuentra muy por encima de su umbral de activación, no debería importar cuánto lo está. Las funciones de score individuales son combinadas mediante la siguiente fórmula, que también se satura a altos (malos) valores:

$$Score = \left[\frac{\sum_{i=1}^{\#x} T(x_i)}{1 + \sum_{i=1}^{\#x} T(x_i)} \right]$$

⁴En nuestro modelo de estudio, las especies RNA **mgta**, **pbgp** y **pmrd**, aunque el esquema es válido para cualquier especie genérica. Las dos primeras indican activación del sistema PhoP/PhoQ y PmrA/PmrB, respectivamente, y la tercera indica si el sistema PhoP/PhoQ se encuentra ejerciendo un estímulo sobre el sistema PmrA/PmrB.

2.4. Marco Conceptual para la Formulación del Algoritmo

En esta sección se describen los conceptos utilizados para inferir la o las redes genéticas asociadas a un problema biológico experimental, tomando como ejemplo la red genética generada por la interacción de los sistemas de dos componentes PmrA/PmrB y PhoP/PhoQ presentes en *Salmonella enterica*. Estos conceptos resumen un mecanismo de decisión adaptativo, que permite generar, validar o refutar, y regenerar hipótesis biológicas sobre el funcionamiento del sistema:

Formalizar un problema biológico particular en base a un modelo continuo que permita representar la dinámica del sistema.

Caracterizar el problema, describiendo al sistema en función de sus componentes funcionales.

Modelizar el problema generando una solución válida para el mismo.

Contrastar propiedades sistémicas emergentes del modelo, tales como su robustez o su flexibilidad.

Posibilitar, en base al resultado de los puntos anteriores,

- la **Predicción** de características funcionales y estructurales del sistema.
- la **Decisión** de si una arquitectura de red genética determinada es factible como solución al problema en estudio.
- la **Adaptación** del modelo en función a la información acumulada, generando nuevas hipótesis acerca de la arquitectura del sistema.

Validar experimentalmente mediante pruebas de laboratorio, la verosimilitud del modelo obtenido, mediante técnicas tales como Chromatin Immunoprecipitation (ChIP) o microarrays específicos.

Mediante esta aproximación al problema se desarrolla un modelo que permite validar computacionalmente las hipótesis derivadas de los resultados experimentales formalizados, y asimismo generar nuevas hipótesis biológicas basadas en características observadas en los modelos obtenidos.

2.4.1. Realismo

Con el fin de testear la capacidad de las arquitecturas estudiadas para reproducir el comportamiento real de la red genética en el organismo vivo

se realizan sucesivas pruebas. El modelo general para las mismas consiste primeramente en, dada una cierta arquitectura, proponer concentraciones iniciales para las especies que modelen las utilizadas en los experimentos biológicos. Luego se realizan iterativamente asignaciones aleatorias de valores a los parámetros dentro de los rangos de valores biológicamente válidos (en base a lo propuesto en [vDMMO00a]), seleccionando aquellos conjuntos que al simular el comportamiento de la red reproducen las funcionalidades pedidas (llamaremos a este conjunto “solución”).

Las funcionalidades, o constraints, a cumplir por una arquitectura pueden ejemplificarse como se muestra en al tabla 2.2.

#Parámetros	<i>NodoEntrada₁</i>	<i>NodoEntrada₂</i>	<i>NodoSalida₁</i>	Proporción	Probabilidad
62	1	0	1	0,00010728	0,86293171

Cuadro 2.2: Ejemplo de funcionalidad a reproducir por una arquitectura. *NodoEntrada_x* indica un nodo distinguido como de entrada para la arquitectura. *NodoSalida_x* indica un nodo distinguido como de salida para la arquitectura. Los valores utilizados para la misma son binarios, indicándose con 1 que el nodo se encuentra activado, y con 0 el caso contrario. Dado que los valores utilizados para modelar las concentraciones de los nodos son reales entre 0 y 1, por activado debe entenderse que el nodo alcanza un valor mayor a un cierto umbral de activación en un cierto tiempo de la simulación, y por no activado que no logra alcanzar ese valor. Los valores de los nodos de entrada son fijados en el tiempo T_0 de la simulación, mientras que los valores de los nodos de salida son lo que deben alcanzarse en un cierto tiempo T_n de la simulación para considerar que esta es exitosa. Esta funcionalidad puede interpretarse como “si el nodo *NodoEntrada₁* se encuentra activado y el nodo *NodoEntrada₂* se encuentra no activado al comenzar la simulación, durante el transcurso de la misma debe evidenciarse una activación del nodo *NodoSalida₁*”. Traducido a lenguaje biológico, puede interpretarse como “la activación del gen *NodoEntrada₁* y la inactivación del gen *NodoEntrada₂* deben determinar la activación del gen *NodoSalida₁*”

Para ejecutar la función que determina si una cierta arquitectura es realista para un cierto conjunto de funcionalidades a reproducir, resulta necesario definir ciertas constantes numéricas, las cuales serán utilizadas también en las funciones de las secciones subsiguientes. Estas son:

maxIter: Cantidad máxima de iteraciones a realizar por el algoritmo; asegura la finalización.

cantIntentos: Cantidad de iteraciones para testear cada funcionalidad.

propAceptacion: proporción esperable de soluciones válidas para una cierta funcionalidad.

En el pseudocódigo utilizado se modela la corrida de una simulación mediante la función `correrSimulacion(arq,p,func)` que devuelve el valor booleano "Verdadero" si la simulación realizada sobre la arquitectura `arq` con los parámetros de las ecuaciones diferenciales instanciados con los valores contenidos en el conjunto `p` cumple con la funcionalidad (constraint) `func`, y "Falso" en caso contrario. La determinación de si una arquitectura, con un cierto conjunto de valores de los parámetros, cumple con las funcionalidades pedidas, se realiza de la siguiente manera:

INPUT:

`arq:` Arquitectura a testear.

`funcs:` Conjunto de funcionalidades a testear para la arquitectura `arq`.

OUTPUT:

`ret:` Valor booleano que determina si la arquitectura cumple con las funcionalidades `funcs`.

`verificarRealismo(arq, funcs)`

`ret = True`

`for f in funcs do:`

`cant = 0`

`for 1 to cantIntentos do:`

`p <- generarParametros(arq)`

`//True si la arq con los`

`//parametros p cumple la funcionalidad f.`

`cumple <- correrSimulacion(arq, p, f)`

`if (cumple) then:`

`cant++`

`else:`

`skip`

`ret <- ret & (cant / cantIntentos >= propAceptacion)`

`return ret`

Este esquema de búsqueda permite calcular la frecuencia de soluciones válidas obtenidas, y en base a esta calcular la probabilidad de encontrar por azar un conjunto que reproduzca el patrón pedido. Si bien en el algoritmo sólo se utiliza la frecuencia, una alta probabilidad de encontrar aleatoriamente conjuntos de parámetros que hagan que la arquitectura cumpla con las funcionalidades pedidas puede indicar que esta funcionalidad es más dependiente de la arquitectura en si misma que de los parámetros escogidos. Si consideramos que p es probabilidad de escoger aleatoriamente un valor para un parámetro que confiera a la arquitectura la habilidad de reproducir una funcionalidad, f la frecuencia de soluciones encontradas y n la cantidad de parámetros, se cumple que:

$$p^n = f$$

de esta manera, la probabilidad de encontrar aleatoriamente un conjunto de parámetros que constituyan una solución válida⁵ para la arquitectura puede calcularse como:

$$p = 10^{\frac{\log f}{n}}$$

Al realizar las pruebas de realismo sobre una arquitectura, se considera a la red genética que esta modela como una “caja negra” a la cual se pide que reproduzca una cierta funcionalidad⁶ en una cierta proporción de los conjuntos de parámetros generados aleatoriamente sin plantearse requerimientos adicionales acerca de otras funcionalidades que la arquitectura podría reproducir o no. El objetivo de las mismas es sólomente determinar si la arquitectura es capaz de reproducir una cierta funcionalidad. Como puede verse en el código de la función utilizada, los conjuntos de parámetros utilizados para cada simulación son distintos, por lo que en este punto no puede determinarse si la arquitectura puede cumplir todas las funcionalidades planteadas *con un mismo conjunto de parámetros*.

Si bien este esquema podría parecer insuficiente para un problema genérico, los resultados obtenidos en [vDMMO00b], [MvDMO02a] y [MvDMO02b] sugieren que para un problema como el estudiado la noción de soluciones globalmente óptimas carece de sentido biológico, y muestran una distribución de las soluciones en el espacio de búsqueda que disminuye significativamente la utilidad de algoritmos de búsqueda más refinados. La posibilidad de

⁵Un conjunto de parámetros es una solución válida para la arquitectura se al realizar una simulación sobre la misma con esos parámetros esta produce los valores requeridos por una cierta funcionalidad, o constraint.

⁶Se requiere que la arquitectura, al asignar ciertos valores de concentración a las especies distinguidas como “de entrada” en el tiempo T_0 , obtenga valores de concentración específicos en los nodos de “salida” en un cierto tiempo de la simulación T_n posterior.

determinar el realismo de una arquitectura propuesta independientemente de los parámetros escogidos constituye una gran ventaja sobre los enfoques de modelización basados en ecuaciones diferenciales como el propuesto por Batchelor y Goulian en [BG03a] para el sistema de dos componentes EnvZ/OmpR, el cual depende de un único conjunto de valores para los parámetros de la arquitectura.

2.4.2. Flexibilidad y Completitud Funcional

Para evaluar la flexibilidad de las arquitecturas, se realizan nuevas búsquedas utilizando como criterio para aceptar una solución la conjunción de todas las condiciones propuestas. Esto es posible realizando simulaciones independientes para cada conjunto de parámetros generado, cada una conteniendo distintos valores iniciales de concentraciones de las especies. Se simulan así las distintas combinaciones de señales a las que se ve expuesta la red real. El cálculo de si la arquitectura es flexible se realiza mediante la siguiente función:

INPUT:

```

  arq: Arquitectura a testear.
  funcs: Conjunto de funcionalidades a testear para la
  arquitectura arq.

```

OUTPUT:

```

  ret: Valor booleano que determina si la arquitectura
  cumple con la conjunción de las funcionalidades funcs.
  verificarFlexibilidad(arq, funcs)
  cant = 0
  for 1 to cantIntentos do:
    p <- generarParametros(arq)
    cumple = True
    for f in funcs do:
      cumple <- cumple & correrSimulacion(arq, p, f)
    if (cumple) then:
      cant++
    else:
      skip

  ret <- cant / cantIntentos >= propAceptacion
  return ret

```

Un conjunto de parámetros es considerado una solución válida cuando los resultados de las simulaciones reproducen todas las funcionalidades requeridas. Se considera entonces que una arquitectura es flexible en su funcionalidad en base a la proporción de conjuntos valores de los parámetros

generados aleatoriamente que le permiten reproducir todas las funcionalidades propuestas.

La determinación de las funcionalidades a cumplir no es en absoluto trivial, dado que estas deben cubrir todas las combinaciones de patrones de entrada - salida a las cuales la red genética modelada debe responder en el organismo real. Por ejemplo, debe asegurarse que las funcionalidades modelen el hecho de que si no se somete a la red a ningún estímulo, esta no debería reproducir ningún comportamiento específico, salvo que este sea el caso en la red genética estudiada. De esta manera, puede proponerse que una red que cumple con la propuesta flexibilidad funcional es necesariamente *completa* en su funcionalidad. Esta definición de flexibilidad funcional de una arquitectura, asociada al concepto de completitud funcional, constituye una gran ventaja sobre el enfoque propuesto para la modelización basada en ecuaciones diferenciales propuesto por von Dassow et All en [vDMMO00b] y [MvDMO02a], en el cual la flexibilidad de las arquitecturas obtenidas no se analiza en forma exhaustiva.

En base a este procedimiento, es posible distinguir el subconjunto de arquitecturas “flexibles funcionalmente” dentro del conjunto de arquitecturas “realistas”.

2.4.3. Robustez de los Parámetros

Una vez definida una medida para el realismo y la flexibilidad de las redes propuestas, surge el interrogante de si estas propiedades son dependientes de los parámetros específicos escogidos, o si son realmente intrínsecas a la topología, como se sugiere en las secciones anteriores. Cabe preguntar entonces si, dada una solución válida para el sistema, esta deja de cumplir con el comportamiento pedido al modificar alguno de sus parámetros. Y en el caso de que una perturbación modifique el comportamiento, puede preguntarse cuánto es necesario perturbar este parámetro.

Para obtener una medida de la robustez, se toma una solución escogida aleatoriamente entre las que reproducen el la conjunción de todos los constraints, y se calcula la puntuación obtenida por la red al modificar un parámetro particular de una solución (y dejando los demás fijos), haciendo que este tome valores a lo largo de todo el intervalo definido como biológicamente significativo. Este proceso se repite para todos los parámetros. Se considera que la red reproduce el patrón requerido si obtiene un score no mayor a una cierta puntuación umbral.

Para la formulación de la función que determina la robustez de la arquitectura con respecto a los parámetros, se adicionan a las referidas en la sección 2.4.1 las siguientes constantes:

cantPuntos: Cantidad de puntos a tomar dentro del rango valido de valores para un parametro de la red.

propRobustezParam: al analizar el poder de la arquitectura para reproducir una funcionalidad dada, si esta se mantiene para esta proporción de los valores analizados el parámetro es considerado robusto.

propParamsRobustos: la solución es considerada robusta con respecto a sus ejes si la proporción de parámetros robustos iguala o supera esta cantidad.

propSolucionesRob: la arquitectura es considerada robusta con respecto a sus ejes si la proporción de soluciones con parámetros robustos iguala o supera esta cantidad.

La determinación de la robustez con respecto a los parámetros de la arquitectura se realiza mediante la función que se expone a continuación:

INPUT:

arq: Arquitectura sobre la cual verificar robustez.
 funcs: Conjunto de funcionalidades a cumplir por la arquitectura arq.

OUTPUT:

ret: Booleano que determina si la arquitectura es robusta con respecto a sus ejes.

```
robustezEjes(arq, funcs)
  soluciones <- soluciones(arq) //conjunto de conjuntos de parametros
                                //que son soluciones para la arq
                                //calculado en un paso anterior
  cantParams <- cantParams(arq) //cantidad de parametros en arq
  cantSolucionesRob <- 0
  ret <- True
  for sol in soluciones do:
    cantParamsRob <- 0
    for p in sol do:
      cantPuntRob <- 0
      //generar una serie de puntos equidistantes dentro
      //del rango valido para el parametro p
      serie <- generarSeriePuntos(p, cantPuntos)
      for punto in serie do:
        // asignar el valor punto al parametro p en el conjunto
        // sol
        sol' <- setParam(sol, p, punto)
        esRob <- correrSimulacion(arq, sol', func)
        if esRob then: cantPuntRob++ else: skip
      if cantPuntRob / cantPuntos >= propRobustezParam then:
        cantParamsRob++
    else:
```

```

    skip
  if (cantParamsRob / cantParams >= propParamsRobustos) then:
    cantSolucionesRob++
  else:
    skip

ret <- cantSolucionesRob / size(soluciones) >= propSolucionesRob
return ret

```

2.4.4. Robustez de las Concentraciones Iniciales

La modelización efectuada sobre el problema biológico presenta limitaciones con respecto a los valores que se asignan a las especies del modelo, debido a que estos son fijados al comenzar la simulación mientras que en el sistema biológico una uniformidad tan estricta no parece razonable. En el caso de los valores de los parámetros de las ecuaciones diferenciales, esta problemática ya ha sido abordada. Sin embargo, resta analizar la robustez de las arquitecturas propuesta con respecto a los valores iniciales de las concentraciones de las especies.

Para medir el impacto de la variación de las concentraciones iniciales en la capacidad funcional de la red, se realizan simulaciones sobre soluciones obtenidas en los pasos anteriores, a las cuales se les han introducido cambios en los valores iniciales de las especies (nodos de la red). Luego se calcula la proporción de estas que conserva la funcionalidad deseada luego de las perturbaciones, considerándose que la arquitectura es robusta si esta proporción es mayor a un cierto umbral.

Para la formulación de la función que determina la robustez de la arquitectura con respecto a las concentraciones iniciales, se adiciona a las referidas en la sección 2.4.1 anterior la siguiente constante:

propSolsRobustasNodos: proporción de soluciones robustas al variar los valores iniciales de los nodos a partir de la cual se considera robusta a la arquitectura.

La robustez de la arquitectura con respecto a los valores de los nodos se computa utilizando la siguiente función:

```

INPUT:
  arq: Arquitectura sobre la cual verificar robustez.
  funcs: Conjunto de funcionalidades a cumplir por la
  arquitectura arq.
OUTPUT:
  ret: Booleano que determina si la arquitectura es robusta
  con respecto a sus nodos.
robustezNodos(arq, funcs)

```

```

soluciones <- soluciones(arq) //conjunto de conjuntos de parametros
                             //que son solucion para la arquitectura
                             //calculado en un paso anterior
//Generar un conjunto de conjuntos de valores iniciales
//para los nodos.
condIniciales <- generarCondInic(arq)
cantSolsRob <- 0

for sol in soluciones do:
  for cond in condIniciales do:
    arq' <- setCondInic(arq, cond)
    esRob <- correrSimulacion(arq', sol, func)
    if esRob then: cantSolsRob++ else: skip

ret <- (cantSolsRob / (#soluciones * #condIniciales))
      >= propSolsRobustasNodos
return ret

```

Es importante destacar que en todos los casos se trata de modificaciones “suaves” sobre el sistema. Efectivamente, el problema biológico estudiado corresponde a un sistema de transducción de señales, y por lo tanto no resultaría coherente esperar un comportamiento uniforme de la red al modificar excesivamente los valores que determinan la presencia o no de estas señales y de las respuestas del sistema.

2.5. Algoritmo GENIE

Este algoritmo constituye una condensación de los métodos utilizados durante el estudio del problema, y constituye tanto la formalización del esquema de trabajo como un resultado de las experiencias realizadas durante el mismo. Para la definición de las distintas etapas, se analizaron diferentes alternativas computacionales, siendo las presentadas la más representativas. Una vez realizada esta formalización, queda abierta la puerta para que, en trabajos futuros, se optimicen las componentes funcionales del algoritmo en base a la experiencia adquirida.

2.5.1. Diagrama de Flujo

El flujo de la ejecución del algoritmo GENIE puede resumirse mediante el diagrama de flujo que se presenta en esta sección. En este se resumen los principales procesos involucrados en la generación, testeo y regeneración de las arquitecturas propuestas para modelar la red genética estudiada. En los distintos puntos de control del diagrama se muestran las etapas del algoritmo en las que se realizan las pruebas de propiedades enunciadas en las secciones

anteriores. Desde una óptica de mayor nivel, GENIE puede interpretarse como una metodología para el análisis y modelado de redes genéticas, cuyos componentes son descriptos en este capítulo, siendo el diagrama de flujo de la figura 2.9 un resumen de la misma.

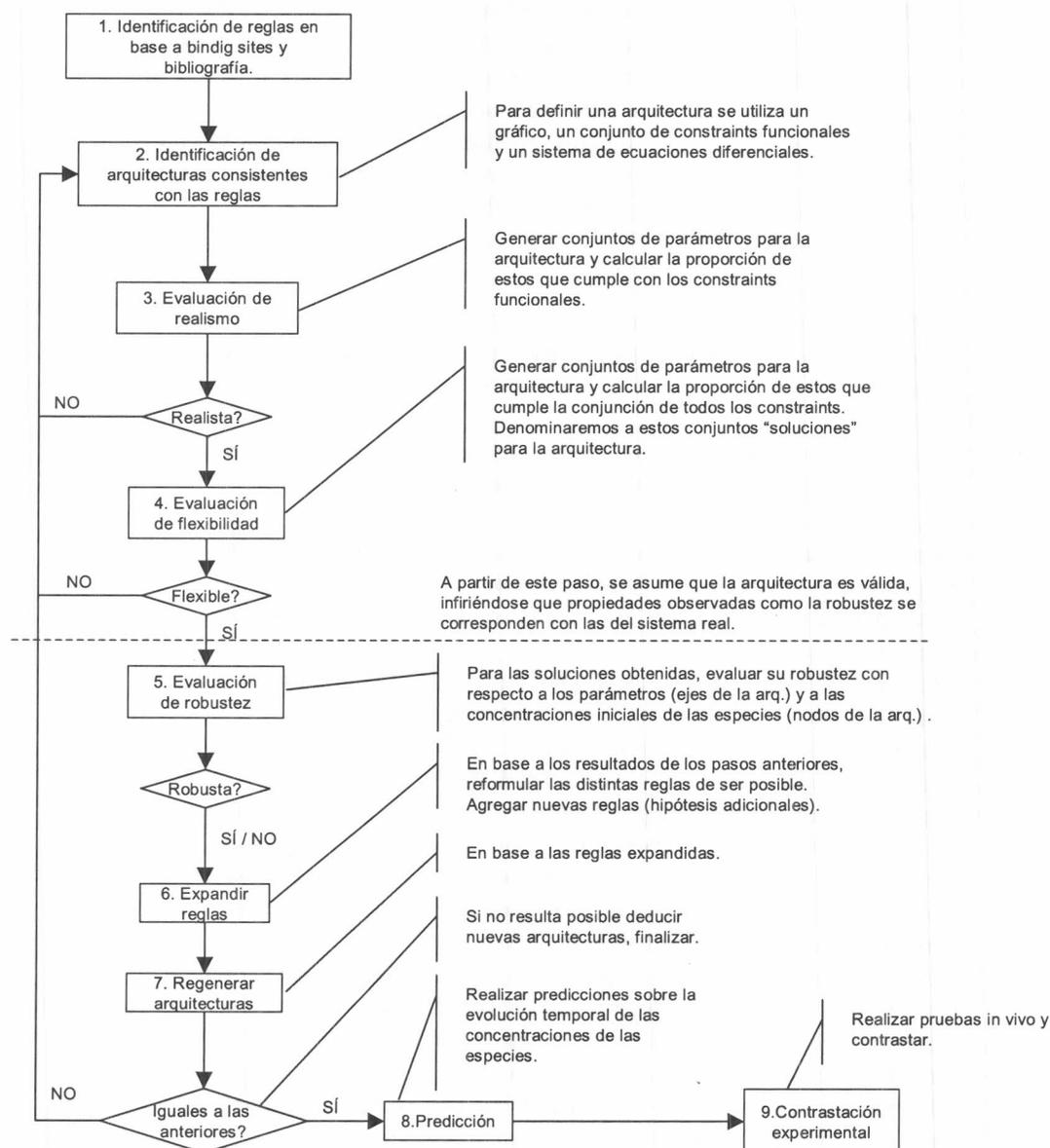


Figura 2.9: Diagrama de flujo para el algoritmo GENIE.

La numeración de los distintos procesos del diagrama se corresponde con la utilizada en el pseudocódigo del algoritmo para identificar a las distintas funciones.

2.5.2. Algoritmo

A continuación se presenta el algoritmo GENIE, que formaliza los distintos procesos involucrados en el diagrama de flujo anteriormente expuesto, definiendo las distintas pruebas numéricas efectuadas. En la función principal del algoritmo se encuentran numerados los llamados a función, indicando el proceso del diagrama con el cual se corresponden. Para la identificación de todas las arquitecturas consistentes con un conjunto de reglas, se utiliza un esquema de inferencia basado en estas últimas, un conjunto de restricciones que filtra las arquitecturas trivialmente inválidas y una librería de módulos que implementan las reglas. Estas arquitecturas son luego testeadas en base a distintas propiedades, como realismo, flexibilidad y robustez.

INPUT:

reglas: Conjunto de reglas que definen al sistema.
funcs: Conjunto de funcionalidades a reproducir por el sistema.

OUTPUT:

arqsValidas: conjunto de arquitecturas factibles para el sistema (realistas, robustas y flexibles).

GLOBALS:

maxIter: Cantidad maxima de iteraciones a realizar por el algoritmo, asegura la finalizacion.
cantIntentos: Cantidad de iteraciones para testear cada funcionalidad.
propAceptacion: proporcion esperable de soluciones validas para una cierta funcionalidad.
propRobustezParam: al analizar el poder de la arquitectura para reproducir una funcionalidad dada, si esta se mantiene para esta proporcion de los valores analizados el parametro es considerado robusto.
propParamsRobustos: la solucion es considerada robusta con respecto a sus ejes si la proporcion de parametros robustos iguala o supera esta cantidad.
propSolucionesRob: la arquitectura es considerada robusta con respecto a sus ejes si la proporcion de soluciones con parametros robustos iguala o supera esta cantidad.
propSolsRobustasNodos: proporcion de soluciones robustas al variar los valores iniciales de los nodos a partir de la cual se considera robusta a la arquitectura.
cantPuntos: Cantidad de puntos a tomar dentro del rango valido de valores para un parametro de la red.

GENIE(reglas, funcs)

```

iter = 1
arqsValidas <- {}
arqsCandidatas <- generarArquitecturas(reglas) (2)
arqsAnt <- {}
while(iter <= maxIter) do:
  arqsRealistas <- {}
  for arq in arqsCandidatas do:
    esRealista <-
      verificarRealismo(arq, funcs) (3)
    if (esRealista) then:
      add(arqsRealistas, arq)
    else:
      skip
  for arq in arqsRealistas do:
    esFlexible <-
      verificarFlexibilidad(arq, funcs) (4)
    if esFlexible then:
      rEjes <- robustezEjes(arq, conjuncion(funcs)) (5)
      rNodos <- robustezNodos(arq, conjuncion(funcs)) (5)
      if (rEjes && rNodos) then:
        add(arqsValidas, arq)
      else:
        skip

  reglas <- expandirReglas(arqsValidas, reglas) (6)
  arqsAnt <- arqsCandidatas
  arqsCandidatas <- generarArquitecturas(reglas) (7)
  if arqsAnt == arqsCandidatas then: break else: skip
  iter++

return arqsValidas

INPUT:
  arq: Arquitectura a testear.
  funcs: Conjunto de funcionalidades a testear para la
  arquitectura arq.
OUTPUT:
  ret: Valor booleano que determina si al arquitectura
  cumple con las funcionalidades funcs.
verificarRealismo(arq, funcs)
  ret = True
  for f in funcs do:
    cant = 0
    for 1 to cantIntentos do:

```

```

    p <- generarParametros(arq)
    //True si la arq con los
    //parametros p cumple la funcionalidad f.
    cumple <- correrSimulacion(arq, p, f)
    if (cumple) then:
        cant++
    else:
        skip
    ret <- ret & (cant / cantIntentos >= propAceptacion)

```

```

return ret

```

INPUT:

```

    arq: Arquitectura a testear.
    funcs: Conjunto de funcionalidades a testear para la
    arquitectura arq.

```

OUTPUT:

```

    ret: Valor booleano que determina si al arquitectura
    cumple con la conjuncion de las funcionalidades funcs.
verificarFlexibilidad(arq, funcs)
    cant = 0
    for 1 to cantIntentos do:
        p <- generarParametros(arq)
        cumple = True
        for f in funcs do:
            cumple <- cumple & correrSimulacion(arq, p, f)
        if (cumple) then:
            agregarParams(arq,p) //agregar los parametros p al
                                //conj. de soluciones de la
                                //arquitectura arq

            cant++
        else:
            skip

    ret <- cant / cantIntentos >= propAceptacion
return ret

```

INPUT:

```

    arq: Arquitectura sobre la cual verificar robustez.
    funcs: Conjunto de funcionalidades a cumplir por la
    arquitectura arq.

```

OUTPUT:

```

    ret: Booleano que determina si la arquitectura es robusta

```

```

con respecto a sus ejes.
robustezEjes(arq, funcs)
  soluciones <- soluciones(arq) //conjunto de conjuntos de parametros
                                //que son soluciones para la arq
  cantParams <- cantParams(arq) //cantidad de parametros en arq
  cantSolucionesRob <- 0
  ret <- True
  for sol in soluciones do:
    cantParamsRob <- 0
    for p in sol do:
      cantPuntRob <- 0
      //generar una serie de puntos equidistantes dentro
      //del rango valido para el parametro p
      serie <- generarSeriePuntos(p, cantPuntos)
      for punto in serie do:
        // asignar el valor punto al parametro p en el conjunto
        // sol
        sol' <- setParam(sol, p, punto)
        esRob <- correrSimulacion(arq, sol', func)
        if esRob then: cantPuntRob++ else: skip
      if cantPuntRob / cantPuntos >= propRobustezParam then:
        cantParamsRob++
      else:
        skip
    if (cantParamsRob / cantParams >= propParamsRobustos) then:
      cantSolucionesRob++
    else:
      skip

  ret <- cantSolucionesRob / size(soluciones) >= propSolucionesRob
  return ret

```

INPUT:

arq: Arquitectura sobre la cual verificar robustez.
 funcs: Conjunto de funcionalidades a cumplir por la
 arquitectura arq.

OUTPUT:

ret: Booleano que determina si la arquitectura es robusta
 con respecto a sus nodos.

```

robustezNodos(arq, funcs)
  soluciones <- soluciones(arq) //conjunto de conjuntos de parametros
  //Generar un conjunto de condiciones iniciales
  //para los valores de los nodos.

```

```

condIniciales <- generarCondInic(arq)
cantSolsRob <- 0

for params in soluciones do:
  for cond in condIniciales do:
    arq' <- setCondInic(arq, cond)
    esRob <- correrSimulacion(arq', params, func)
    if esRob then: cantSolsRob++ else: skip

ret <- (cantSolsRob / (#soluciones * #condIniciales))
  >= propSolsRobustasNodos
return ret

```

INPUT:

arq: arquitectura para la cual se genera un conjunto de parametros.

OUTPUT:

params: conjunto de valores de los parametros para la arquitectura arq.

generarParametros(arq)

Asignar valores para cada uno de los parametros de la arquitectura arq de acuerdo al algoritmo utilizado, e.g. aleatoriamente.

INPUT:

arq: Arquitectura sobre la cual se realiza la simulacion.
p: Conjunto de valores para los parametros de la arquitectura.
func: Funcionalidad a reproducir en la simulacion.

OUTPUT:

ret: Booleano determinando si la simulacion genero los resultados esperados de acuerdo a func.

correrSimulacion(arq, p, func)

Realizar la simulacion de acuerdo a lo descrito en el texto y en la bibliografia.

INPUT:

reglas: Conjunto de reglas que definen al sistema.

OUTPUT:

arqs: Conjunto de arquitecturas que cumplen las reglas especificadas.

generarArquitecturas(reglas)

Generar el conjunto de arquitecturas que se ajustan a las reglas especificadas, de acuerdo a lo descrito en el texto y a la bibliografía.

INPUT:

reglas: Conjunto de reglas que definen al sistema.

arqsValidas: conjunto de arquitecturas validas obtenidas hasta la presente iteracion.

OUTPUT:

reglas': Conjunto de reglas expandidas que definen al sistema.

expandirReglas(arqsValidas, reglas)

En base a las arquitecturas obtenidas hasta el momento como validas y a las reglas especificadas, agregar reglas (hipotesis) que permitan profundizar la descripcion del sistema obtenida hasta el momento en funcion de la evidencia acumulada (arquitecturas en arqsValidas).

2.6. Observaciones Finales

En este capítulo se enunciaron los elementos que permiten formular la metodología GENIE para el análisis de redes genéticas. Esta permite transformar la evidencia de los binding sites de los genes de la red genética modelada y la información existente en la bibliografía en un conjunto de reglas que determinan el comportamiento de la red. Estas reglas son luego modeladas mediante módulos arquitectónicos existentes en una librería de módulos, combinándose en una arquitectura completa. Esta arquitectura es implementada realizando un mapeo entre los módulos y las ecuaciones diferenciales existentes en una librería de ecuaciones diferenciales. El sistema de ecuaciones diferenciales obtenido permite realizar simulaciones sobre el sistema, testeando propiedades tales como su realismo, flexibilidad y robustez. La validación o no de estas propiedades aporta importante información para la inferencia de nuevas hipótesis sobre el comportamiento del sistema. Este esquema iterativo de formulación, prueba y reformulación puede repetirse hasta obtener una arquitectura válida para el sistema estudiado, con el fin de realizar predicciones sobre el mismo. Estas predicciones permitirán dirigir la experimentación biológica, acelerando el proceso de exploración de las redes genéticas estudiadas.

Capítulo 3

Aplicación de GENIE a la Red Genética Constituida por la Interacción de los Sistemas de Dos Componentes PhoP/PhoQ y PmrA/PmrB

3.1. Introducción

En este capítulo se presenta la aplicación de GENIE a la red genética constituida por la interacción de los sistemas de dos componentes PhoP/PhoQ y PmrA/PmrB presentes en *Salmonella enterica serovar Typhimurium*. El sistema de dos componentes PmrA/PmrB (regulador de respuesta/sensor, respectivamente) responde independientemente a dos señales, alto nivel de Fe³⁺ extracelular, sensado por PmrB, y bajo nivel de Mg²⁺, sensado por la proteína PhoQ. El sistema PhoP/PhoQ (regulador de respuesta/sensor, respectivamente) constituye un regulador maestro que gobierna la adaptación a medios de bajo Mg²⁺ y la virulencia en ratones así como también otras funciones biológicas. Su funcionamiento como sistema de dos componentes es en principio análogo al anterior, siendo el estímulo activador el bajo nivel de Mg²⁺ extracelular. Entre los genes activados por PhoP se encuentra el gen *pmrD*, el cual resulta de especial interés porque presenta un sitio de binding para PmrA, y su producto, la proteína PmrD, puede asociarse a la proteína PmrA. Ambos sistemas de dos componentes muestran una coordinación de sus funciones in vivo, aunque los mecanismos exactos de interacción son aún desconocidos.

La aplicación de GENIE a este problema tiene como puntos principales la determinación de las reglas que definen el comportamiento de la red, la formulación y reformulación iterativa de sus arquitecturas asociadas, el testeo de propiedades de estas últimas como el realismo, flexibilidad y robustez, y la realización de predicciones y la contrastación experimental.

En la sección 3.2 se explica cómo la evidencia de binding sites para los genes de la red genética estudiada permite determinar las principales reglas que determinan la interacción entre las especies modeladas.

En las secciones 3.3.1 a 3.3.8 se referencia a las iteraciones del algoritmo, describiendo las arquitecturas obtenidas en base a su diagrama de arquitectura, sus constraints a reproducir y su sistema de ecuaciones diferenciales, y los resultados numéricos de las pruebas realizadas sobre las mismas. Con fines puramente descriptivos, se asigna un nombre coloquial a las arquitecturas.

Para cada iteración, se presentan los resultados numéricos obtenidos durante la ejecución del algoritmo, en cuanto al realismo y la flexibilidad de la arquitectura, determinando si las arquitecturas responden al comportamiento denominado "real" para el sistema, en el sentido de modelar las tuplas estímulo-respuesta definidas en base al comportamiento observado en *Salmonella*. Estos resultados permitieron reformular en distintas ocasiones las arquitecturas, refinando adaptativamente los modelos hasta alcanzar un comportamiento aceptablemente descriptivo del sistema. Asimismo, facilitaron la formulación de hipótesis alternativas sobre los mecanismos biológicos involucrados. En el caso de resultar las pruebas satisfactorias (sección 3.3.8), se muestran también los resultados para la robustez de la arquitectura. Estos determinan el carácter robusto o no de las soluciones obtenidas (conjuntos de valores de los parámetros de la red), tanto en lo referente a los valores de los parámetros como a los valores de las concentraciones iniciales de las distintas especies.

En la sección 3.3.1 se describe un modelo inicial basado en la literatura existente para el problema [KLG03][HS95] y la evidencia experimental de sitios de binding (binding-sites) más una serie de hipótesis genéricas adicionales. Luego, en base a la información aportada por la validación de propiedades y a las nuevas hipótesis obtenidas, se describen los distintos modelos generados. Aquí se muestra el poder del método para explorar el problema formalizado, generando modelos alternativos basados en las nuevas hipótesis biológicas obtenidas.

Luego de exponer los resultados puntuales obtenidos para cada iteración, en la sección 3.4 se muestra un resumen de todos los resultados obtenidos, incluyendo gráficas comparativas entre las distintas iteraciones.

Finalmente, en la sección 3.5 se realizan predicciones sobre el modelo y se compara la similitud entre las series temporales obtenidas en las simulaciones y las obtenidas experimentalmente in vivo.

3.2. Modelado de Interacciones Genéticas Basado en Binding Sites

Con el fin de aprender modelos para los diferentes binding sites para factores de transcripción se agruparon ejemplos de la base de datos RegulonDB, los cuales fueron prototipados utilizando modelos de weight matrix. Luego, se realizó una búsqueda en las regiones intergénicas de *Salmonella* utilizando estos modelos, detectándose la co-ocurrencia de distintos motivos de binding sites para factores de transcripción y PhoP “boxes” putativas dentro de las mismas regiones intergénicas. La distancia entre la PhoP box y los otros binding sites putativos fueron agrupadas. Las mencionadas distancias fueron prototipadas utilizando funciones de fuzzy membership y utilizadas para caracterizar la relación entre las PhoP boxes putativas y otras binding-boxes mediante la evaluación de su distancia de acuerdo con casos previamente conocidos reportados en bases de datos. Se utilizó la herramienta SOAR [ZTG03] para identificar interacciones de regulación involucrando a la proteína PhoP. En la Figura 3.1 se observan ejemplos de distintas regiones promotoras encontradas. En la Figura 3.2 se muestra el esquema de interacción entre los genes de los subsistemas PhoP/PhoQ y PmrA/PmrB derivado de la información obtenida del estudio de los binding sites.

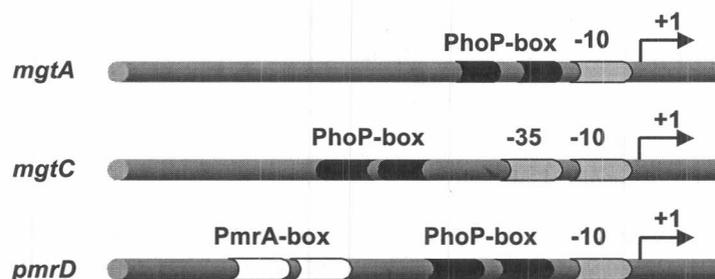


Figura 3.1: Regiones promotoras de los genes *mgtA*, *mgtC* y *PmrD*.

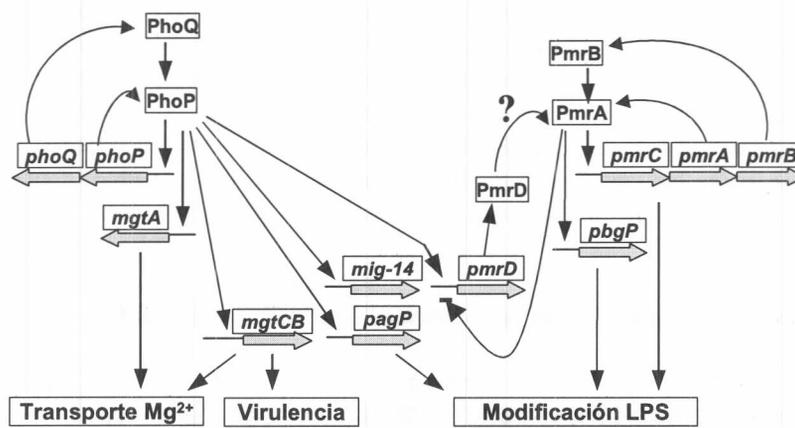


Figura 3.2: Interacciones entre los genes de los subsistemas PhoP/PhoQ y PmrA/PmrB derivadas de la información obtenida del estudio de los binding sites.

3.3. Resultados Obtenidos de la Aplicación de GENIE a la Red Genética PhoP/PhoQ -PmrA/PmrB

Las sucesivas iteraciones del algoritmo GENIE produjeron distintas arquitecturas para la red genética estudiada. En el Cuadro 3.1 puede verse un resumen de las arquitecturas obtenidas.

44 **CAPÍTULO 3. APLICACIÓN DE GENIE A UNA RED GENÉTICA**

Iteración I	Nombre Reducido	Feedback + PhoP / PmrA	Feedback - PmrA-P / PmrA-P_PmrD
II	Reducido sin Feedback	PhoP / PmrA	No
III	Alternativa 1	PhoP / PmrA	PmrA-P / PmrA-P_PmrD
IV	Alternativa 2.1	PhoP-P / PmrA-P	PmrA-P / PmrA-P_PmrD
V	Alternativa 2.2	PhoP-P / PmrA-P	No
VI	Alternativa 2.3	sin autoreg. phoP-phoQ / con autoreg. vía PmrA-P	PmrA-P / PmrA-P_PmrD
VII	Alternativa 2.4	PhoP-P / PmrA-P	PmrA-P / PmrA-P_PmrD
VIII	Enriquecido	PhoP-P / PmrA-P	PmrA-P / PmrA-P_PmrD
Iteración I	Crosstalk No	Activación pbgP PmrA-P y PmrA-P_PmrD	Fosforilación PhoQ-ACT fosforila a PhoP y defosforila a PhoP-P. PmrB-ACT fosforila a PmrA y defosforila a PmrA-P. PhoP-P y PmrA-P se defosforilan espontáneamente.
II	No	PmrA-P y PmrA-P_PmrD	PhoQ-ACT fosforila a PhoP y defosforila a PhoP-P. PmrB-ACT fosforila a PmrA y defosforila a PmrA-P. PhoP-P y PmrA-P se defosforilan espontáneamente.
III	Sí	PmrA-P y PmrA-P_PmrD. PhoP-P basal al comenzar el experimento.	PhoQ-ACT fosforila a PhoP y defosforila a PhoP-P. PmrB-ACT fosforila a PmrA y defosforila a PmrA-P. PhoP-P y PmrA-P se defosforilan espontáneamente.
IV	Sí	PmrA-P y PmrA-P_PmrD	PhoQ-ACT fosforila a PhoP y defosforila a PhoP-P. PmrB-ACT fosforila a PmrA y defosforila a PmrA-P. PhoP-P y PmrA-P se defosforilan espontáneamente.
V	Sí	PmrA-P y PmrA-P_PmrD	PhoQ-ACT fosforila a PhoP y defosforila a PhoP-P. PmrB-ACT fosforila a PmrA y defosforila a PmrA-P. PhoP-P y PmrA-P se defosforilan espontáneamente.
VI	Sí	PmrA-P y PmrA-P_PmrD	PhoQ-ACT fosforila a PhoP y defosforila a PhoP-P. PmrB-ACT fosforila a PmrA y defosforila a PmrA-P. PhoP-P y PmrA-P se defosforilan espontáneamente.
VII	Sí	PmrA-P y PmrA-P_PmrD activan a pbgP con distinta probabilidad fija, sin permitirse que varíen estos dos parámetros. PhoP-P basal al comenzar el experimento.	PhoQ-ACT fosforila a PhoP y defosforila a PhoP-P. PmrB-ACT fosforila a PmrA y defosforila a PmrA-P. PhoP-P y PmrA-P se defosforilan espontáneamente.
VIII	Sí	PmrA-P y PmrA-P_PmrD	PhoQ fosforila a PhoP y defosforila a PhoP-P. PmrB fosforila a PmrA y defosforila a PmrA-P. PhoQ-ACT fosforila a PhoP y PmrB-ACT fosforila a PmrA. PhoP-P y PmrA-P se defosforilan espontáneamente.

Cuadro 3.1: Resumen de las arquitecturas obtenidas para la red genética PhoP/PhoQ - PmrA/PmrB.

Las arquitecturas se encuentran clasificadas de acuerdo a las principales características que las distinguen. **Feedback +** indica qué especies se proponen como activadoras de los operones phoQ-phoP y pmrB_pmrA, respectivamente. **Feedback -** indica qué especies se proponen como inhibidoras del gen pmrD. **Fosforilación** indica la actividad de las proteínas sensoras PhoQ y PmrB sobre PhoP y PmrA, respectivamente. **Activación pbgP** indica si se propone alguna variación para la activación de este gen por parte de PmrA-P y PmrA-P_PmrD. **Crosstalk** indica si se requiere que la arquitectura presente un crosstalk entre los dos subsistemas, evidenciado como una activación de pbgP en presencia de bajo Mg²⁺ y bajo Fe³⁺.

3.3.1. Iteración 1: Reducido

En base a la combinación de la evidencia acumulada mediante el estudio de los binding sites de los genes de *Salmonella* y de las características básicas del sistema presentes en la literatura, puede realizarse una primera aproximación simple al modelado de la red genética constituida por los sistemas PmrA/PmrB y PhoP/PhoQ, a la cual denominaremos modelo Reducido. Este modelo condensa la información aportada por las primeras reglas utilizadas; en los casos en los que no se poseía información acerca de una cierta interacción, se utilizó la hipótesis más genérica posible. Luego, las sucesivas iteraciones sobre la arquitectura permitieron formular hipótesis alternativas para refinar estas elecciones, dando lugar a nuevos modelos.

Este primer modelo se desprende de las reglas enunciadas a continuación. El sistema PhoP/PhoQ es activado en bajo nivel de Mg²⁺ extracelular, y el sistema PmrA/PmrB es activado en alto nivel de Fe³⁺ extracelular. Estas condiciones son censadas por las proteínas de membrana PhoQ y PmrB, respectivamente, las cuales pasan en presencia de la señal a un estado “activado” en el cual las especies actúan como kinasas y fosfatasa de sus reguladores de respuesta asociados e.g. PhoP y PmrA, respectivamente. Las formas fosfatadas de estas últimas dos especies funcionan como activadoras de la transcripción de distintos genes, tomando en nuestro ejemplo a *mgfA* para PhoP y *pbpP* para PmrA. Las proteínas PhoP y PhoQ son traducidas de un mismo RNA *phop_phoq*, constituyendo un operón. Análogamente, PmrA y PmrB son traducidas de un mismo RNA *pmra_pmrB*. Estos dos operones son regulados positivamente por PhoP y PmrA, respectivamente.

Con respecto a la interconexión entre ambos sistemas de dos componentes, consideraremos la existencia de una conexión de “ida” desde el subsistema PhoP/PhoQ al subsistema PmrA/PmrB, y una conexión de “vuelta” en sentido inverso. La evidencia de binding sites y de experimentos de CHIP permite afirmar que la expresión de *pmrD* es inhibida por PmrA, siendo esta la conexión de vuelta entre los sistemas. Dado que la evidencia de binding sites permite afirmar que no existe una interacción entre el producto del gen *pmrD* y el DNA del gen PmrA, que constituiría una conexión de ida entre los sistemas entre los sistemas a nivel transcripcional, se asume que la interacción entre *pmrD* y PmrA es post-transcripcional. Esta interacción consiste en que la forma fosfatada de PmrA es protegida por PmrD de la actividad fosfatasa de PmrB.

Descripción de la Arquitectura

En la Fig. 3.3 puede apreciarse un esquema del modelo Reducido. Con respecto a las convenciones de nombres utilizadas para las especies, pueden darse las siguientes definiciones: se utilizan letras mayúsculas para las especies proteicas, minúsculas para los RNA, el sufijo -P para las especies fos-

foriladas y el sufijo -ACT para identificar especies proteicas que sufren un cambio que modifica su actividad. Las especies que son ácidos ribonucleicos (RNA), son representadas con círculos, mientras que las especies proteicas son representadas con rectángulos.

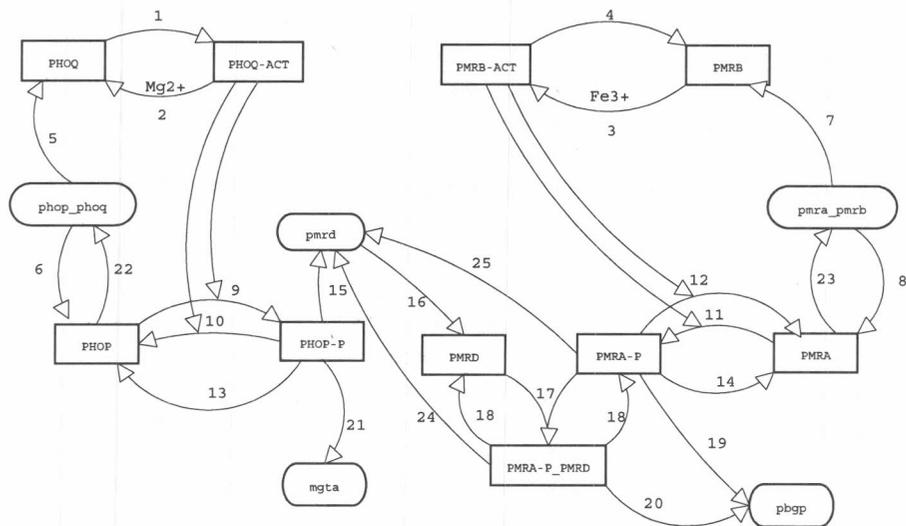


Figura 3.3: Modelo Reducido.

- 1 Bajo nivel de Mg^{2+} favorece el estado PHOQ-ACT (IVADO) en el equilibrio
- 2 Alto nivel de Mg^{2+} favorece el estado PHOQ en el equilibrio
- 3 Alto nivel de Fe^{3+} favorece el estado PMRB-ACT (IVADO) en el equilibrio
- 4 Bajo nivel de Fe^{3+} favorece el estado PMRB en el equilibrio
- 5 phop_phoq es traducido a PHOQ
- 6 phop_phoq es traducido a PHOP
- 7 pmra_pmr es traducido a PMRB
- 8 pmra_pmr es traducido a PMRA
- 9 PHOP es fosforilado a PHOP-P via PHOQ-ACT actividad kinasa
- 10 PHOP-P es defosforilado a PHOP via PHOQ-ACT actividad fosfatasa
- 11 PMRA es fosforilado a PMRA-P via PMRB-ACT actividad kinasa
- 12 PMRA-P es defosforilado a PMRA via PMRB-ACT actividad fosfatasa
- 13 PHOP-P es defosforilado a PHOP via defosforilación espontánea
- 14 PMRA-P es defosforilado a PMRA via defosforilación espontánea
- 15 PHOP-P activa la transcripción de pmrd
- 16 pmrd es traducido a PMRD
- 17 PMRD se asocia con PMRA-P para constituir la especie PMRA-P_PMRD que activa la expresión de pbgp e inhibe la de pmrd pero no es afectada por la actividad fosfatasa de PMRB-ACT
- 18 PMRA-P_PMRD se disocia para formar PMRD y PMRA-P
- 19 PMRA-P activa la transcripción de pbgp
- 20 PMRA-P_PMRD activa la transcripción de pbgp
- 21 PHOP-P activa la transcripción de mgta
- 22 PHOP activa la transcripción de phop_phoq
- 23 PMRA activa la transcripción de pmra_pmr
- 24 PMRA-P_PMRD inhibe la transcripción de pmrd
- 25 PMRA-P inhibe la transcripción de pmrd

3.3. RESULTADOS OBTENIDOS DE LA APLICACIÓN DE GENIE 47

De acuerdo a la bibliografía analizada, puede inferirse un primer patrón de actividad (o “salida” del modelo) para el sistema en función de distintos estímulos, o “entradas”. Estos patrones se muestran en el Cuadro 3.2 mediante valores booleanos (1 ó 0) por simplicidad¹, debiendo considerarse que un 1 indica activado, y un 0 inactivado. En el caso del Mg, un 1 indica la presencia del estímulo (bajo Mg en el caso estudiado) y un 0 el caso contrario. En el Fe, un 1 implica la presencia de estímulo (es decir alto Fe en el sistema estudiado) y un 0 el caso contrario. Los valores de las concentraciones de las especies Fe y Mg corresponden a la “entrada” del sistema, mientras que los valores de mgta, pbgp, y pmrd corresponden a la “salida” del sistema (es decir, la señal transducida). Valores altos de mgta y pbgp indican la activación de los subsistemas phoP/phoQ y pmrA/pmrB, respectivamente. Un valor alto de pmrd indica la activación de la conexión de ida entre los sistemas, y un valor bajo indica la activación de la conexión de vuelta (feedback negativo). En general, los valores indicados en las tablas para las especies de entrada se mantienen constantes a lo largo de las simulaciones, mientras que los de las especies de salida son los esperados al transcurrir la simulación.

Constraint	Mg	Fe	mgta	pmrd	pbgp
1	1	1	1	0	1
2	1	0	1	1	0
3	0	1	0	0	1
4	0	0	0	0	0

Cuadro 3.2: Patrones de entrada-salida para el modelo Reducido.

En el caso del segundo constraint, con Mg activo y Fe inactivo, debe notarse que se pide que el gen pbgP (el cual es activado por el sistema PmrA-PmrB; la activación de pbgP modela en forma genérica la activación de este sistema) no se encuentre activado. La información analizada hasta este punto del análisis referente a la actividad post-transcripcional de PMRD no permite modelar un mecanismo que activaría pbgP en ausencia de PMRA-P, por lo que la activación de pbgP depende de la presencia del estímulo de Fe³⁺ en este modelo.

Otro punto de interés consiste en que la transcripción de los operones phoP-phoQ y pmrA_pmrB depende de las especies defosforiladas de PhoP y PmrA. Esta es una hipótesis conservadora a la luz de la evidencia de binding, que no distingue entre una y otra especie. Este punto será analizado nuevamente en secciones subsiguientes.

¹En todos los experimentos, se utilizaron valores reales para las concentraciones de las especies, y se consideraron distintos umbrales de activación. Sin embargo, se presentan los patrones en función de especies activadas o no activadas para facilitar el análisis conceptual de las funcionalidades

Ecuaciones diferenciales para el modelo Reducido

Los superíndices en los términos de las ecuaciones indican la reacción del Cuadro 3.21 y de la Figura 3.3 de la cual se derivan.

Especie	Adición	Sustracción
phop_phoq	22	
PHOQ	2, 5	1
PHOQ-ACT	1	2
PHOP	6, 10,13	9
PHOP-P	9	10, 13
mgta	21	
pmra_pmr	23	
PMRB	4, 7	3
PMRB-ACT	3	4
PMRA	8, 12, 14	11
PMRA-P	11, 18	12, 14, 17
pbgp	19, 20	
PMRD	16, 18	17
PMRA-P_PMRD	17	18
pmrd	15	24,25

Cuadro 3.3: Términos de las ecuaciones que determinan las concentraciones de las especies.

Adición: Reacciones que determinan un término sustracción en la ecuación diferencial de la especie.

Sustracción: Reacciones que determinan un término de sustracción en la ecuación diferencial de la especie..

Decaimiento: A todas la especies se les agrega un término de sustracción extra para modelar el decaimiento de primer orden.

La numeración se corresponde con la utilizada para numerar los ejes en la Figura 3.3.

3.3. RESULTADOS OBTENIDOS DE LA APLICACIÓN DE GENIE 49

$$\begin{aligned}
 \frac{d[\text{PHOQ}]}{dt} &= T_0 \left(\frac{1}{H_{\text{PHOQ}}} [\text{phop-phoq}]^{(5)} + r_{\text{PHOQACTmax_Mg2}} [\text{PHOQACT}] [\text{Mg2}]^{(2)} \right. \\
 &\quad \left. - [\text{PHOQ}] r_{\text{PHOQ2PHOQACT}}^{(1)} - \frac{[\text{PHOQ}]^{(\text{dec})}}{H_{\text{PHOQ}}} \right) \\
 \frac{d[\text{phop-phoq}]}{dt} &= T_0 \left(\frac{1}{H_{\text{phop-phoq}}} \frac{[\text{PHOP}]^{\nu_{\text{PHOP-phop-phoq}}}}{K_{\text{PHOP-phop-phoq}}^{\nu_{\text{PHOP-phop-phoq}}} + [\text{PHOP}]^{\nu_{\text{PHOP-phop-phoq}}}} \right)^{(22)} - \frac{[\text{phop-phoq}]^{(\text{dec})}}{H_{\text{phop-phoq}}} \\
 \frac{d[\text{PHOQ_ACT}]}{dt} &= T_0 \left([\text{PHOQ}] r_{\text{PHOQ2PHOQACT}}^{(1)} - r_{\text{PHOQACTmax_EMg2}} [\text{PHOQACT}] [\text{EMg2}]^{(2)} \right. \\
 &\quad \left. - \frac{[\text{PHOQACT}]^{(\text{dec})}}{H_{\text{PHOQACT}}} \right) \\
 \frac{d[\text{PHOP}]}{dt} &= T_0 \left(\frac{1}{H_{\text{PHOP}}} [\text{phop-phoq}]^{(6)} \right. \\
 &\quad + P_{\text{PHOQACTP}} [\text{PHOPP}] \frac{[\text{PHOQACT}]^{\nu_{\text{PHOQACT_PHOPP}}}}{K_{\text{PHOQACT_PHOPP}}^{\nu_{\text{PHOQACT_PHOPP}}} + [\text{PHOQACT}]^{\nu_{\text{PHOQACT_PHOPP}}}} \right)^{(10)} \\
 &\quad + r_{\text{PHOPP}} [\text{PHOPP}]^{(13)} \\
 &\quad - P_{\text{PHOQACTK}} [\text{PHOP}] \frac{[\text{PHOQACT}]^{\nu_{\text{PHOQACT_PHOP}}}}{K_{\text{PHOQACT_PHOP}}^{\nu_{\text{PHOQACT_PHOP}}} + [\text{PHOQACT}]^{\nu_{\text{PHOQACT_PHOP}}}} \right)^{(9)} \\
 &\quad - \frac{[\text{PHOP}]^{(\text{dec})}}{H_{\text{PHOP}}} \\
 \frac{d[\text{PHOPP}]}{dt} &= T_0 \left(P_{\text{PHOQACTK}} [\text{PHOP}] \frac{[\text{PHOQACT}]^{\nu_{\text{PHOQACT_PHOP}}}}{K_{\text{PHOQACT_PHOP}}^{\nu_{\text{PHOQACT_PHOP}}} + [\text{PHOQACT}]^{\nu_{\text{PHOQACT_PHOP}}}} \right)^{(9)} \\
 &\quad - P_{\text{PHOQACTP}} [\text{PHOPP}] \frac{[\text{PHOQACT}]^{\nu_{\text{PHOQACT_PHOPP}}}}{K_{\text{PHOQACT_PHOPP}}^{\nu_{\text{PHOQACT_PHOPP}}} + [\text{PHOQACT}]^{\nu_{\text{PHOQACT_PHOPP}}}} \right)^{(10)} \\
 &\quad - [\text{PHOPP}] r_{\text{PHOPP}}^{(13)} - \frac{[\text{PHOPP}]^{(\text{dec})}}{H_{\text{PHOPP}}} \\
 \frac{d[\text{mgta}]}{dt} &= T_0 \left(\frac{1}{H_{\text{mgta}}} \frac{[\text{PHOPP}]^{\nu_{\text{PHOPP_mgta}}}}{K_{\text{PHOPP_mgta}}^{\nu_{\text{PHOPP_mgta}}} + [\text{PHOPP}]^{\nu_{\text{PHOPP_mgta}}}} \right)^{(21)} - \frac{[\text{mgta}]^{(\text{dec})}}{H_{\text{mgta}}} \\
 \frac{d[\text{pmra_pmrb}]}{dt} &= T_0 \left(\frac{1}{H_{\text{pmra_pmrb}}} \frac{[\text{PMRAP}]^{\nu_{\text{PMRAP_pmra_pmrb}}}}{K_{\text{PMRAP_pmra_pmrb}}^{\nu_{\text{PMRAP_pmra_pmrb}}} + [\text{PMRAP}]^{\nu_{\text{PMRAP_pmra_pmrb}}}} \right)^{(23)} \\
 &\quad - \frac{[\text{pmra_pmrb}]^{(\text{dec})}}{H_{\text{pmra_pmrb}}} \\
 \frac{d[\text{PMRB}]}{dt} &= T_0 \left(\frac{1}{H_{\text{PMRB}}} [\text{pmrd}]^{(7)} + r_{\text{PMRBACT}} [\text{PMRBACT}]^{(4)} \right. \\
 &\quad \left. - r_{\text{PMRBACTmax_Fe3}} [\text{PMRBACT}] [\text{Fe3+}]^{(3)} - \frac{[\text{PMRB}]^{(\text{dec})}}{H_{\text{PMRB}}} \right) \\
 \frac{d[\text{PMRBACT}]}{dt} &= T_0 \left(r_{\text{PMRBACTmax_Fe3}} [\text{Fe3+}]^{(3)} + [\text{PMRB}] [\text{Fe3+}]^{(3)} - [\text{PMRBACT}] r_{\text{PMRBACT}}^{(4)} \right. \\
 &\quad \left. - \frac{[\text{PMRB}]^{(\text{dec})}}{H_{\text{PMRB}}} \right) \\
 \frac{d[\text{PMRA}]}{dt} &= T_0 \left(\frac{1}{H_{\text{PMRA}}} [\text{pmra_pmrb}]^{(8)} + r_{\text{PMRAP}} [\text{PMRAP}]^{(14)} \right. \\
 &\quad + P_{\text{PMRBACTP}} [\text{PMRAP}] \frac{[\text{PMRBACT}]^{\nu_{\text{PMRBACT_PMRA}}}}{K_{\text{PMRBACT_PMRA}}^{\nu_{\text{PMRBACT_PMRA}}} + [\text{PMRBACT}]^{\nu_{\text{PMRBACT_PMRA}}}} \right)^{(12)} \\
 &\quad - P_{\text{PMRBACTK}} [\text{PMRA}] \frac{[\text{PMRBACT}]^{\nu_{\text{PMRBACT_PMRAP}}}}{K_{\text{PMRBACT_PMRAP}}^{\nu_{\text{PMRBACT_PMRAP}}} + [\text{PMRBACT}]^{\nu_{\text{PMRBACT_PMRAP}}}} \right)^{(11)} \\
 &\quad - \frac{[\text{PMRA}]^{(\text{dec})}}{H_{\text{PMRA}}} \\
 \frac{d[\text{PMRAP}]}{dt} &= T_0 \left(P_{\text{PMRBACTP}} [\text{PMRA}] \frac{[\text{PMRBACT}]^{\nu_{\text{PMRBACT_PMRA}}}}{K_{\text{PMRBACT_PMRA}}^{\nu_{\text{PMRBACT_PMRA}}} + [\text{PMRBACT}]^{\nu_{\text{PMRBACT_PMRA}}}} \right)^{(11)} \\
 &\quad + \frac{[\text{PMRAP_PMRD}]^{(18)}}{H_{\text{PMRAP_PMRD}}}
 \end{aligned}$$

50 CAPÍTULO 3. APLICACIÓN DE GENIE A UNA RED GENÉTICA

$$\begin{aligned}
 & -P_{PMRBACT}K[PMRAP] \frac{[PMRBACT]^{\nu_{PMRBACT-PMRAP}}}{K_{PMRBACT-PMRAP}^{\nu_{PMRBACT-PMRAP}} + [PMRBACT]^{\nu_{PMRBACT-PMRAP}}} \quad (12) \\
 & -r_{PMRAP}[PMRAP]^{(14)} - r_{PMRAP-PMRDmax-PMRD}[PMRAP][PMRD]^{(17)} \\
 & - \frac{[PMRAP]^{(dec)}}{H_{PMRAP}} \\
 \frac{d[pbgp]}{dt} = & T_0 \left(\frac{1}{H_{pbgp}} (1 - (1 - \alpha_{PMRAP} \frac{1}{H_{pbgp}} \frac{[PMRAP]^{\nu_{PMRAP-pbgp}}}{K_{PMRAP-pbgp}^{\nu_{PMRAP-pbgp}} + [PMRAP]^{\nu_{PMRAP-pbgp}}}) \right. \\
 & (1 - \alpha_{PMRAP-PMRD} \frac{1}{H_{pbgp}} \\
 & \left. \frac{[PMRAP-PMRD]^{\nu_{PMRAP-PMRD-pbgp}}}{K_{PMRAP-PMRD-pbgp}^{\nu_{PMRAP-PMRD-pbgp}} + [PMRAP-PMRD]^{\nu_{PMRAP-PMRD-pbgp}}}) \right) \quad (19,20) \\
 & - \frac{[pbgp]^{(dec)}}{H_{pbgp}} \\
 \frac{d[PMRD]}{dt} = & T_0 \left(\frac{1}{H_{PMRD}} [pmrd]^{(16)} - r_{PMRAP-PMRDmax-PMRD}[PMRAP][PMRD]^{(17)} \right. \\
 & \left. - \frac{[PMRD]^{(dec)}}{H_{PMRD}} \right) \\
 \frac{d[PMRAP-PMRD]}{dt} = & T_0 (r_{PMRAP-PMRDmax-PMRD}[PMRAP][PMRD]^{(17)} \\
 & - \frac{[PMRAP-PMRD]^{(18)}}{H_{PMRAP-PMRD}}) \\
 \frac{d[pmrd]}{dt} = & T_0 \left(\frac{1}{H_{pmrd}} \frac{[PHOPP]^{\nu_{PHOPP-pmrd}}}{K_{PHOPP-pmrd}^{\nu_{PHOPP-pmrd}} + [PHOPP]^{\nu_{PHOPP-pmrd}}} \right. \\
 & (1 - \frac{[PMRAP]^{\nu_{PMRAP-pmrd}}}{K_{PMRAP-pmrd}^{\nu_{PMRAP-pmrd}} + [PMRAP]^{\nu_{PMRAP-pmrd}}}) \\
 & \left. (1 - \frac{[PMRAP-PMRD]^{\nu_{PMRAP-PMRD-pmrd}}}{K_{PMRAP-PMRD-pmrd}^{\nu_{PMRAP-PMRD-pmrd}} + [PMRAP-PMRD]^{\nu_{PMRAP-PMRD-pmrd}}}) \right) \quad (15,24,25) - \frac{[pmrd]^{(dec)}}{H_{pmrd}}
 \end{aligned}$$

Resultados Numéricos: Realismo y Flexibilidad

Con el fin de testear la capacidad de las arquitecturas estudiadas para reproducir el comportamiento real de la red se realizaron sucesivas pruebas. El modelo general para las mismas consistió primeramente en, dada una cierta arquitectura, proponer concentraciones iniciales para las especies que modelen las utilizadas en los experimentos biológicos (incluyendo las especies extracelulares Mg²⁺ y Fe²⁺, cuyos valores constituyen el “patrón de entrada” del modelo). Luego se asignaron rangos de valores biológicamente válidos a los distintos parámetros del modelo y se realizaron iterativamente caminatas aleatorias por el conjunto de parámetros (espacio de solución), seleccionando aquellos conjuntos que al simular el comportamiento de la red reprodujeron los patrones pedidos (llamaremos a este conjunto “solución”). Este esquema de búsqueda permite calcular la frecuencia de soluciones válidas obtenidas, y en base a esta calcular la probabilidad de encontrar por azar un conjunto que reproduzca el patrón pedido. Los resultados de estas pruebas se muestran en el Cuadro 3.4, en el cual se evidencia que la arquitectura es realista con respecto a los constraints planteados. Puede observarse que el constraint 4 resulta más sencillo de reproducir que los demás, lo cual es coherente dado que se trata del “blanco” de la prueba, es decir en el cual no debe evidenciarse actividad al suprimir ambos estímulos. Se observan diferencias de al menos un orden de magnitud entre las proporciones obtenidas para cada uno de los otros constraints, pudiendo ordenarse a estos por orden decreciente de “facilidad” de la arquitectura para reproducirlos en el orden 2, 3, 1.

Constraint	#Parámetros	Mg	Fe	mgta	pmrd	pbgp	Proporción	Probabilidad
1	60	1	1	1	0	1	0,00035474	0,875988521
2	60	1	0	1	1	0	0,06442205	0,955323753
3	60	0	1	0	0	1	0,0019469	0,901202504
4	60	0	0	0	0	0	0,66514196	0,993227127

Cuadro 3.4: Resultados para el modelo Reducido.

Constraint: número de la condición a cumplir, de acuerdo a lo descrito en la sección de modelos descubiertos. **#Parámetros:** cantidad de parámetros presentes en las ecuaciones del modelo. **Mg, Fe:** nodos cuyo valor equivale al estímulo de entrada para la red, representándose esquemáticamente con 0 y 1 la ausencia y presencia de estímulo, respectivamente. **mgta, pbgp, pmrd:** nodos cuyo valor equivale a la salida de la red, representándose con 1 para mgta y pbgp la activación de los sistemas PhoP/PhoQ y PmrA/PmrB, respectivamente, y con 1 para pmrd la influencia del sistema PhoP/PhoQ en PmrA/PmrB; 0 representa no actividad en todos los casos. **Proporción:** proporción de conjuntos de parámetros que cumplen con el constraint requerido del total de conjuntos generados. **Probabilidad:** calculada en función de la proporción, probabilidad de que al seleccionar un parámetro al azar para la red este le permita cumplir exitosamente con el constraint.

Al analizar la Flexibilidad de la arquitectura (Fig. 3.5), se determina que esta presenta soluciones capaces de reproducir todos los patrones de comportamiento requeridos en forma individual. Es decir, estas soluciones pueden reproducir las distintas combinaciones de señales a las que se ve

expuesta la red real.

Modelo	#Parámetros	Mg	Fe	mgta	pmrd	pbgp	Proporción	Probabilidad
Reducido	60	1 1 0 0	1 0 1 0	1 1 0 0	0 1 0 0	1 0 1 0	1,87764E-05	0,834116778

Cuadro 3.5: Resultados para la conjunción de los constraints 1 & 2 & 3 & 4 para el el modelo Reducido.

Para cada especie, se muestran las cuatro condiciones requeridas. Obsérvese que para un modelo particular, las primeras posiciones del array de valores para cada especie se refieren al constraint 1 expuesto anteriormente, y así sucesivamente.

3.3.2. Iteración 2: Reducido sin Feedback

Descripción de la Arquitectura

Una variación de interés al modelo anteriormente expuesto consiste en no considerar el denominado eje “de vuelta” entre los dos sistemas de dos componentes, reduciendo el acoplamiento entre los mismos (denominaremos a este modelo Reducido Sin Feedback). Para visualizar el modelo obtenido, basta eliminar los ejes etiquetados con los números 24 y 25 en la Fig.3.3, de acuerdo a lo que se muestra en la Fig. 3.4. En el Cuadro 3.6 se muestran los patrones de entrada-salida esperados para considerar válida a una solución.

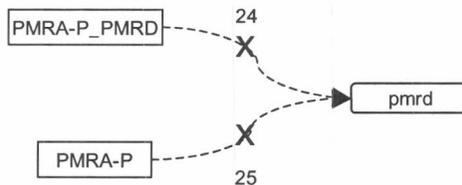


Figura 3.4: Modificaciones al esquema de la figura para el modelo Reducido que determinan la arquitectura del modelo Reducido sin Feedback.

Ecuaciones diferenciales modificadas con respecto al modelo Reducido:

$$\frac{d[\text{pmrd}]}{dt} = T_0 \left(\frac{1}{H_{\text{pmrd}} K_{\text{PHOPP-pmrd}}^{\nu_{\text{PHOPP-pmrd}}} + [\text{PHOPP}]^{\nu_{\text{PHOPP-pmrd}}}} \right)^{(15)} - \frac{[\text{pmrd}]^{(dec)}}{H_{\text{pmrd}}}$$

Constraint	Mg	Fe	mgta	pmrd	pbgp
1	1	1	1	1	1
2	1	0	1	1	0
3	0	1	0	0	1
4	0	0	0	0	0

Cuadro 3.6: Patrones de entrada-salida para el modelo Reducido Sin Feedback.

Nótese que en el caso 1 se requiere que pmrd se encuentre activado, consistentemente con la eliminación de la acción inhibitoria de PMRA.

Resultados Numéricos: Realismo y Flexibilidad

En el Cuadro 3.7 se observan los resultados de las pruebas de Realismo para esta arquitectura. Las proporciones obtenidas para cada constraint permiten ordenar a estos decrecientemente como 4,2,3,1. Este orden en las proporciones es el mismo que el obtenido en la arquitectura anterior, mostrando que el constraint 1 es el más difícil de reproducir para la red, siendo las proporciones obtenidas cuatro órdenes de magnitud inferiores a las del constraint 4. Al realizar las pruebas de Flexibilidad, como se observa en el Cuadro 3.8, puede notarse sin embargo que la proporción de soluciones obtenidas es superior a la obtenida para el modelo Reducido, lo cual indica que si bien el modelo Reducido sin Feedback presenta una menor facilidad para reproducir cada uno de los constraints, las soluciones obtenidas tienen una mayor probabilidad de reproducir los cuatro constraints simultáneamente.

Constraint	#Parámetros	Mg	Fe	mgta	pmrd	pbgp	Proporción	Probabilidad
1	56	1	1	1	1	1	0,00097532	0,883557379
2	56	1	0	1	1	0	0,12128993	0,963029804
3	56	0	1	0	0	1	0,00321704	0,902589529
4	56	0	0	0	0	0	0,66439315	0,992725141

Constraint	#Parámetros	Mg	Fe	mgta	pmrd	pbgp	Proporción	Probabilidad
1*	56	1	1	1	0	1	0,0003643	0,868155465

Cuadro 3.7: Resultados para el modelo Reducido sin Feedback.

Constraint: número de la condición a cumplir, de acuerdo a lo descrito en la sección de modelos descubiertos. **#Parámetros:** cantidad de parámetros presentes en las ecuaciones del modelo. **Mg, Fe:** nodos cuyo valor equivale al estímulo de entrada para la red, representándose esquemáticamente con 0 y 1 la ausencia y presencia de estímulo, respectivamente. **mgta, pbgp, pmrd:** nodos cuyo valor equivale a la salida de la red, representándose con 1 para mgta y pbgp la activación de los sistemas PhoP/PhoQ y PmrA/PmrB, respectivamente, y con 1 para pmrd la influencia del sistema PhoP/PhoQ en PmrA/PmrB; 0 representa no actividad en todos los casos. **Proporción:** proporción de conjuntos de parámetros que cumplen con el constraint requerido del total de conjuntos generados. **Probabilidad:** calculada en función de la proporción, probabilidad de que al seleccionar un parámetro al azar para la red este le permita cumplir exitosamente con el constraint. Resultados obtenidos para el constraint utilizado en el modelo reducido, es decir requiriendo activación de pmrd ante el estímulo del Fe y el Mg.

Modelo	#Parámetros	Mg	Fe	mgta	pmrd	pbgp	Proporción	Probabilidad
Reducido sin Feedback	56	1 1 0 0	1 0 1 0	1 1 0 0	1 1 0 0	1 0 1 0	3,97356E-05	0,834476142
Reducido sin Feedback*	56	1 1 0 0	1 0 1 0	1 1 0 0	0 1 0 0	1 0 1 0	0	0

Cuadro 3.8: Resultados para la conjunción de los constraints 1 & 2 & 3 & 4 para el modelo Reducido sin Feedback.

Para cada especie, se muestran las cuatro condiciones requeridas. Obsérvese que para un modelo particular, las primeras posiciones del array de valores para cada especie se refieren al constraint 1 expuesto anteriormente, y así sucesivamente.

*Conjunción de constraints iguales a los requeridos para el modelo Reducido.

3.3.3. Iteración 3: Alternativa 1**Descripción de la Arquitectura**

Se considera la posibilidad de que el sistema PmrA/PmrB se active, aumentando consecuentemente la transcripción de pbgp, en presencia del estímulo del Mg (bajo Mg) y en ausencia del de Fe (alto Fe). En base a las ecuaciones del modelo presentado anteriormente, en el cual la acción post-transcripcional de PMRD depende de su asociación a PMRA-P, en lugar de fosfatar directamente a PMRA o de activar su transcripción, para que esto sea posible debe formularse una hipótesis adicional. La más conservadora, no mediando más información, es que debe existir una cierta cantidad de PMRA-P al comenzar el experimento. Denominaremos a este modelo "Alternativa 1", y su esquema y ecuaciones no difieren de los presentados para el modelo Reducido, modificándose sólo la concentración inicial de PMRA-P y la tabla de valores de entrada-salida en la forma descripta. No se realizan cambios en el esquema de la arquitectura y en el sistema de ecuaciones diferenciales. Los patrones de entrada-salida quedan modificados como se ve en la Fig.3.9.

Constraint	Mg	Fe	mgta	pmrd	pbgp
1	1	1	1	1	1
2	1	0	1	1	1
3	0	1	0	0	1
4	0	0	0	0	0

Cuadro 3.9: Patrones de entrada-salida para el modelo Alternativa 1.

Resultados Numéricos: Realismo y Flexibilidad

En los resultados para las pruebas de Realismo del Cuadro 3.10, puede notarse que las proporciones obtenidas difieren en su ordenamiento de las obtenidas para las arquitecturas anteriores. En este caso, el constraint que presenta menos soluciones para la arquitectura es el número 2. Esto es coherente con el hecho de que en este modelo se requiere la existencia de un “crosstalk” entre los sistemas, es decir que el gen *pbgP* se active ante el estímulo de Mg^{2+} , sin mediar el estímulo de Fe^{3+} . Esta mayor dificultad motiva que no se presenten soluciones que permitan validar la Flexibilidad de esta arquitectura, como se muestra en el Cuadro 3.11.

Constraint	#Parámetros	Mg	Fe	mgta	pmrd	pbgp	Proporción	Probabilidad
1	60	1	1	1	0	1	0,0008419	0,8886981
2	60	1	0	1	1	1	0,0001273	0,8611505
3	60	0	1	0	0	1	0,0036405	0,9106524
4	60	0	0	0	0	0	0,6597033	0,9930912

Cuadro 3.10: Resultados para el modelo Alternativa 1.

Constraint: número de la condición a cumplir, de acuerdo a lo descrito en la sección de modelos descubiertos. **#Parámetros:** cantidad de parámetros presentes en las ecuaciones del modelo. **Mg, Fe:** nodos cuyo valor equivale al estímulo de entrada para la red, representándose esquemáticamente con 0 y 1 la ausencia y presencia de estímulo, respectivamente. **mgta, pbgp, pmrd:** nodos cuyo valor equivale a la salida de la red, representándose con 1 para *mgta* y *pbgp* la activación de los sistemas *PhoP/PhoQ* y *PmrA/PmrB*, respectivamente, y con 1 para *pmrd* la influencia del sistema *PhoP/PhoQ* en *PmrA/PmrB*; 0 representa no actividad en todos los casos. **Proporción:** proporción de conjuntos de parámetros que cumplen con el constraint requerido del total de conjuntos generados. **Probabilidad:** calculada en función de la proporción, probabilidad de que al seleccionar un parámetro al azar para la red este le permita cumplir exitosamente con el constraint.

Modelo	#Parámetros	Mg	Fe	mgta	pmrd	pbgp	Proporción	Probabilidad
Alternativa 1	60	1100	1010	1100	0100	1110	0	0

Cuadro 3.11: Resultados para la conjunción de los constraints 1 & 2 & 3 & 4 para el modelo Alternativa 1.

Para cada especie, se muestran las cuatro condiciones requeridas. Obsérvese que para un modelo particular, las primeras posiciones del array de valores para cada especie se refieren al constraint 1 expuesto anteriormente, y así sucesivamente.

3.3.4. Iteración 4: Alternativa 2.1

En iteraciones posteriores a las que dieron lugar las hipótesis (arquitecturas) anteriormente enumeradas, surgieron nuevas combinaciones de interés para el estudio del sistema. Estas se agrupan con la denominación “Alternativa 2.*”, derivándose del modelo reducido en base a distintas variaciones que se describen a continuación.

Descripción de la Arquitectura

La primera variación, presente en todas las arquitecturas Alternativa2.*, consiste en proponer que la autoregulación positiva de los operones *phoP-phoQ* y *pmrA_pmrB* es realizada por las especies fosforiladas de PhoP y PmrA (respectivamente), en lugar de las no fosforiladas. Esto es consistente con la idea de que se requiere una mayor expresión de estas especies en un contexto en el cual la presencia de los estímulos de entrada activa los sistemas de transducción de señales, fosforilando estas dos últimas especies. Si bien es posible que en la célula tanto las especies fosforiladas como no fosforiladas contribuyan con distinta magnitud a esta autoregulación (debido a distintas fuerzas de asociación al DNA), la aproximación descrita corresponde a la asignación de una fuerza 0 (nula) para las especies no fosforiladas y 1 para las fosforiladas.

Una pregunta central para todas las arquitecturas es cómo interactúan PmrD y PmrA-P para lograr la activación del gen *pbgP*. La posibilidad estudiada aquí es que PmrD se asocia a PmrA-P, formando un compuesto estable que no es posible de ser defosforilado por PmrB-ACT; siguiendo este mecanismo, resulta esperable que la red active *pbgP* en ausencia de Mg^{2+} ($Mg = 1$ en las tablas) y ausencia de Fe^{3+} ($Fe = 0$ en las tablas). La pregunta a formular es entonces “Es posible que la acción estabilizadora del PmrD al asociarse al PmrA-P sea suficiente para que éste se mantenga a lo largo del experimento (activando la transcripción de *pbgp*), a pesar de que PmrB no está activado al ser $Fe = 0$?” Para que esto sea posible, es necesario que exista un cierto nivel basal de PmrA-P, y por lo tanto se asigna a éste un valor inicial de 0.2 en lugar de 0.0 en todos los experimentos.

Los patrones de entrada-salida a reproducir por el modelo no se modifican con respecto a los de la iteración anterior (Alternativa 1).

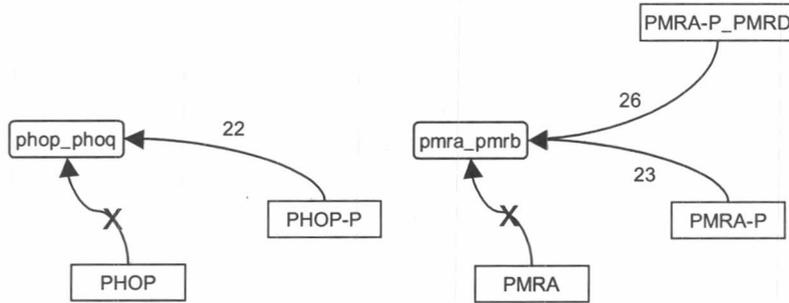


Figura 3.5: Modificaciones al esquema de la figura para el modelo Reducido que determinan la arquitectura del modelo Alternativa 2.1.

Ecuaciones diferenciales modificadas con respecto al modelo Reducido:

Las hipótesis adicionales en esta arquitectura permiten redefinir algunas de las ecuaciones utilizadas en el modelo Reducido, como se muestra a continuación:

$$\frac{d[\text{phop_phoq}]}{dt} = T_0 \left(\frac{1}{H_{\text{phop_phoq}}} \frac{[\text{PHOPP}]^{\nu_{\text{PHOPP_phop_phoq}}}}{K_{\text{PHOPP_phop_phoq}}^{\nu_{\text{PHOPP_phop_phoq}}} + [\text{PHOPP}]^{\nu_{\text{PHOPP_phop_phoq}}}} \right)^{(22)} - \frac{[\text{phop_phoq}]^{(dec)}}{H_{\text{phop_phoq}}}$$

$$\frac{d[\text{pmra_pmrb}]}{dt} = T_0 \left(\frac{1}{H_{\text{pmra_pmrb}}} \left(1 - (1 - \alpha_{\text{PMRAP}}) \frac{1}{H_{\text{pmra_pmrb}}} \frac{[\text{PMRAP}]^{\nu_{\text{PMRAP_pmra_pmrb}}}}{K_{\text{PMRAP_pmra_pmrb}}^{\nu_{\text{PMRAP_pmra_pmrb}}} + [\text{PMRAP}]^{\nu_{\text{PMRAP_pmra_pmrb}}}} \right) \right. \\ \left. (1 - \alpha_{\text{PMRAP_PMRD}}) \frac{1}{H_{\text{pmra_pmrb}}} \frac{[\text{PMRAP_PMRD}]^{\nu_{\text{PMRAP_PMRD_pmra_pmrb}}}}{K_{\text{PMRAP_PMRD_pmra_pmrb}}^{\nu_{\text{PMRAP_PMRD_pmra_pmrb}}} + [\text{PMRAP_PMRD}]^{\nu_{\text{PMRAP_PMRD_pmra_pmrb}}}} \right)^{(23,26)} - \frac{[\text{pmra_pmrb}]^{(dec)}}{H_{\text{pmra_pmrb}}}$$

3.3. RESULTADOS OBTENIDOS DE LA APLICACIÓN DE GENIE 59

Resultados Numéricos: Realismo y Flexibilidad

En el Cuadro 3.12 puede verse que las proporciones obtenidas para los constraints 1, 2 y 3 son similares, siendo equivalentes a las obtenidas para el constraint 2 en el modelo Alternativa 1. Esto indica una disminución en el poder para cumplir con los constraints con respecto a las arquitecturas anteriores. Nuevamente, no se observan soluciones válidas que permitan determinar que esta es una arquitectura Flexible, como se muestra en el Cuadro 3.13.

Constraint	#Parámetros	Mg	Fe	mgta	pmrd	pbgp	Proporción	Probabilidad
Alternativa2.1								
1	62	1	1	1	0	1	0,00010728	0,86293171
2	62	1	0	1	1	1	0,00010331	0,862405877
3	62	0	1	0	0	1	0,00017975	0,870144576
4	62	0	0	0	0	0	0,79079849	0,996221478

Cuadro 3.12: Resultados para el modelo Alternativa 2.1.

Constraint: número de la condición a cumplir, de acuerdo a lo descrito en la sección de modelos descubiertos. **#Parámetros:** cantidad de parámetros presentes en las ecuaciones del modelo. **Mg, Fe:** nodos cuyo valor equivale al estímulo de entrada para la red, representándose esquemáticamente con 0 y 1 la ausencia y presencia de estímulo, respectivamente. **mgta, pbgp, pmrd:** nodos cuyo valor equivale a la salida de la red, representándose con 1 para mgta y pbgp la activación de los sistemas PhoP/PhoQ y PmrA/PmrB, respectivamente, y con 1 para pmrd la influencia del sistema PhoP/PhoQ en PmrA/PmrB; 0 representa no actividad en todos los casos. **Proporción:** proporción de conjuntos de parámetros que cumplen con el constraint requerido del total de conjuntos generados. **Probabilidad:** calculada en función de la proporción, probabilidad de que al seleccionar un parámetro al azar para la red este le permita cumplir exitosamente con el constraint.

Modelo	#Parámetros	Mg	Fe	mgta	pmrd	pbgp	Proporción	Probabilidad
Alternativa 2.1	62	1100	1010	1100	0100	1110	0	0

Cuadro 3.13: Resultados para la conjunción de los constraints 1 & 2 & 3 & 4 para el modelo Alternativa 2.1.

Para cada especie, se muestran las cuatro condiciones requeridas. Obsérvese que para un modelo particular, las primeras posiciones del array de valores para cada especie se refieren al constraint 1 expuesto anteriormente, y así sucesivamente.

3.3.5. Iteración 5: Alternativa 2.2

Descripción de la Arquitectura

En esta arquitectura se modifica la del modelo de la Alternativa 2.1 eliminando el feedback negativo de PmrA y PmrA-P a pmrD, como se ve en la Figura 3.6. No se modifican los patrones de entrada salida con respecto a la iteración anterior.

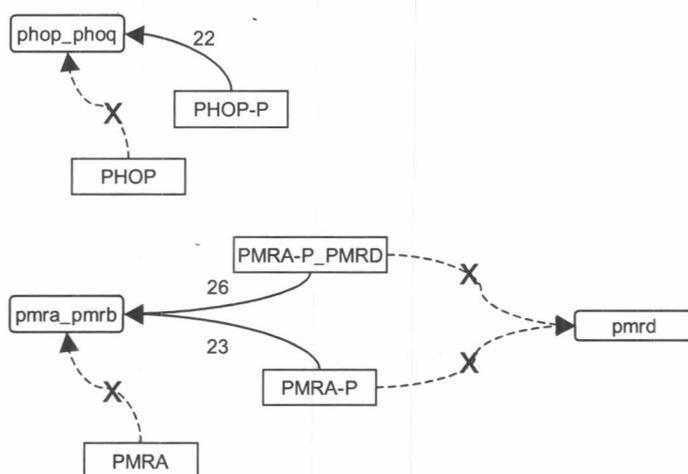


Figura 3.6: Modificaciones al esquema de la figura para el modelo Reducido que determinan la arquitectura del modelo Alternativa 2.2

Ecuaciones diferenciales modificadas con respecto al modelo Reducido:

$$\frac{d[\text{phop-phoq}]}{dt} = T_0 \left(\frac{1}{H_{\text{phop-phoq}}} \frac{[\text{PHOPP}]^{\nu_{\text{PHOPP-phop-phoq}}}}{K_{\text{PHOPP-phop-phoq}}^{\nu_{\text{PHOPP-phop-phoq}}} + [\text{PHOPP}]^{\nu_{\text{PHOPP-phop-phoq}}}} - \frac{[\text{phop-phoq}]^{(\text{dec})}}{H_{\text{phop-phoq}}} \right)^{(22)}$$

$$\frac{d[\text{pmra-pmrB}]}{dt} = T_0 \left(\frac{1}{H_{\text{pmra-pmrB}}} \left(1 - (1 - \alpha_{\text{PMRAP}}) \frac{1}{H_{\text{pmra-pmrB}}} \frac{[\text{PMRAP}]^{\nu_{\text{PMRAP-pmrB}}}}{K_{\text{PMRAP-pmrB}}^{\nu_{\text{PMRAP-pmrB}}} + [\text{PMRAP}]^{\nu_{\text{PMRAP-pmrB}}}} \right) \right. \\ \left. (1 - \alpha_{\text{PMRAP-PMRD}}) \frac{1}{H_{\text{pmra-pmrB}}} \frac{[\text{PMRAP-PMRD}]^{\nu_{\text{PMRAP-PMRD-pmrB}}}}{K_{\text{PMRAP-PMRD-pmrB}}^{\nu_{\text{PMRAP-PMRD-pmrB}}} + [\text{PMRAP-PMRD}]^{\nu_{\text{PMRAP-PMRD-pmrB}}}} \right)^{(23,26)}$$

$$\frac{d[\text{pmrD}]}{dt} = T_0 \left(\frac{1}{H_{\text{pmrD}}} \frac{[\text{PHOPP}]^{\nu_{\text{PHOPP-pmrD}}}}{K_{\text{PHOPP-pmrD}}^{\nu_{\text{PHOPP-pmrD}}} + [\text{PHOPP}]^{\nu_{\text{PHOPP-pmrD}}}} - \frac{[\text{pmrD}]^{(\text{dec})}}{H_{\text{pmrD}}} \right)^{(15)}$$

3.3. RESULTADOS OBTENIDOS DE LA APLICACIÓN DE GENIE 61

Resultados Numéricos: Realismo y Flexibilidad

Al igual que en el modelo Alternativa 2.1, no se observan diferencias en las proporciones de soluciones obtenidas para los constraints 1 a 3. El constraint 4, al igual que en las iteraciones anteriores, presenta una proporción varios órdenes de magnitud superior a la de los demás. Al comparar el Cuadro 3.14 con los Cuadros 3.12 y 3.4 puede observarse que esta proporción tiende a aumentar cuando disminuyen las restantes, lo cual resulta coherente con el hecho de que el constraint 4 requiere que la red no presente actividad alguna ante la falta de estímulo, siendo esto más probable si la arquitectura propuesta posee un menor poder para reproducir las distintas funcionalidades. Como se observa en el Cuadro 3.15, esta arquitectura no presenta soluciones que cumplan los cuatro constraints simultáneamente.

Constraint	#Parámetros	Mg	Fe	mgta	pmrd	pbgp	Proporción	Probabilidad
Alternativa 2.2								
1	58	1	1	1	0	1	8,2774E-05	0,850391367
2	58	1	0	1	1	1	0,00012101	0,855977336
3	58	0	1	0	0	1	0,00046723	0,876149225
4	58	0	0	0	0	0	0,78856455	0,995912841

Cuadro 3.14: Resultados para el modelo Alternativa 2.2.

Constraint: número de la condición a cumplir, de acuerdo a lo descrito en la sección de modelos descubiertos. **#Parámetros:** cantidad de parámetros presentes en las ecuaciones del modelo. **Mg, Fe:** nodos cuyo valor equivale al estímulo de entrada para la red, representándose esquemáticamente con 0 y 1 la ausencia y presencia de estímulo, respectivamente. **mgta, pbgp, pmrd:** nodos cuyo valor equivale a la salida de la red, representándose con 1 para mgta y pbgp la activación de los sistemas PhoP/PhoQ y PmrA/PmrB, respectivamente, y con 1 para pmrd la influencia del sistema PhoP/PhoQ en PmrA/PmrB; 0 representa no actividad en todos los casos. **Proporción:** proporción de conjuntos de parámetros que cumplen con el constraint requerido del total de conjuntos generados. **Probabilidad:** calculada en función de la proporción, probabilidad de que al seleccionar un parámetro al azar para la red este le permita cumplir exitosamente con el constraint.

Modelo	#Parámetros	Mg	Fe	mgta	pmrd	pbgp	Proporción	Probabilidad
Alternativa 2.2	58	1100	1010	1100	0100	1110	0	0

Cuadro 3.15: Resultados para la conjunción de los constraints 1 & 2 & 3 & 4 para el modelo Alternativa 2.2.

Para cada especie, se muestran las cuatro condiciones requeridas. Obsérvese que para un modelo particular, las primeras posiciones del array de valores para cada especie se refieren al constraint 1 expuesto anteriormente, y así sucesivamente.

3.3.6. Iteración 6: Alternativa 2.3

Descripción de la Arquitectura

En Alternativa 2.3 se modifica a la Alternativa 2.1 eliminando la autoregulación de PhoP (tanto fosforilado como no fosforilado) al operón `phoP-phoQ`. De esta manera, no existe activación de `phop-phoq` en este modelo, siendo esta la única modificación a las ecuaciones diferenciales enunciadas para el modelo Alternativa 2.1. No se modifican los patrones de entrada-salida con respecto a la iteración anterior.

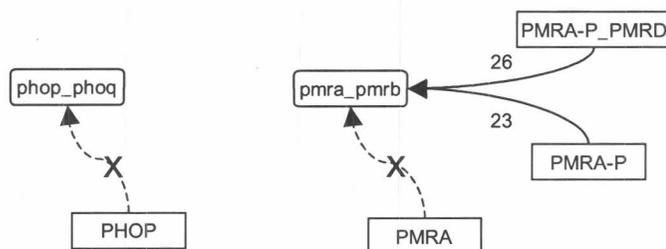


Figura 3.7: Modificaciones al esquema de la figura para el modelo Reducido que determinan la arquitectura del modelo Alternativa 2.3.

3.3. RESULTADOS OBTENIDOS DE LA APLICACIÓN DE GENIE 63

Resultados Numéricos: Realismo y Flexibilidad

Los resultados obtenidos para el Realismo de esta arquitectura, Cuadro 3.16 indican una mayor dificultad para reproducir los constraints 1 y 2, en los cuales está presente el estímulo del Mg2+, con respecto a los constraints 3 y 4. Esto permite inferir que la autoregulación positiva de PhoP-P al operón phoP-phoQ, que se elimina en esta arquitectura, resulta importante para que la red reproduzca los patrones de comportamiento observables en presencia de Mg2+. Estas proporciones son aún inferiores a las obtenidas para las arquitecturas anteriores, y consecuentemente no se observan soluciones para la conjunción de los constraints, como se observa en el Cuadro 3.17.

Constraint	#Parámetros	Mg	Fe	mgta	pmrd	pbgp	Proporción	Probabilidad
Alternativa 2.3								
1	60	1	1	1	0	1	4,8346E-05	0,847369295
2	60	1	0	1	1	1	5,0393E-05	0,84795508
3	60	0	1	0	0	1	0,0005726	0,883006975
4	60	0	0	0	0	0	0,30381418	0,980340175

Cuadro 3.16: Resultados para el modelo Alternativa 2.3.

Constraint: número de la condición a cumplir, de acuerdo a lo descrito en la sección de modelos descubiertos. **#Parámetros:** cantidad de parámetros presentes en las ecuaciones del modelo. **Mg, Fe:** nodos cuyo valor equivale al estímulo de entrada para la red, representándose esquemáticamente con 0 y 1 la ausencia y presencia de estímulo, respectivamente. **mgta, pbgp, pmrd:** nodos cuyo valor equivale a la salida de la red, representándose con 1 para mgta y pbgp la activación de los sistemas PhoP/PhoQ y PmrA/PmrB, respectivamente, y con 1 para pmrd la influencia del sistema PhoP/PhoQ en PmrA/PmrB; 0 representa no actividad en todos los casos. **Proporción:** proporción de conjuntos de parámetros que cumplen con el constraint requerido del total de conjuntos generados. **Probabilidad:** calculada en función de la proporción, probabilidad de que al seleccionar un parámetro al azar para la red este le permita cumplir exitosamente con el constraint.

Modelo	#Parámetros	Mg	Fe	mgta	pmrd	pbgp	Proporción	Probabilidad
Alternativa 2.3	60	1100	1010	1100	0100	1110	0	0

Cuadro 3.17: Resultados para la conjunción de los constraints 1 & 2 & 3 & 4 para el modelo Alternativa 2.3.

Para cada especie, se muestran las cuatro condiciones requeridas. Obsérvese que para un modelo particular, las primeras posiciones del array de valores para cada especie se refieren al constraint 1 expuesto anteriormente, y así sucesivamente.

3.3.7. Iteración 7: Alternativa 2.4

Descripción de la Arquitectura

En Alternativa 2.4 se modifica a Alternativa 2.1 planteándose el siguiente interrogante. Si existe una cierta concentración basal de PmrA-P, y este puede activar a pbgP, cómo se explica que pbgP no se encuentre basalmente activado pero sí se active en presencia de bajo Mg²⁺ (Mg =1) por intermedio de PmrD? Para explorar esta cuestión se agrega el constraint de que la fuerza de activación de pbgP del complejo PmrA-P_PmrD sea mayor que la de PmrA-P sólomente. Esto se logra mediante un modelado probabilístico, proponiendo que la probabilidad de que una molécula de PmrA-P_PmrD se asocie al DNA de pbgP es mayor a la de PmrA-P (por ejemplo, 0.8 y 0.2 respectivamente).

Dado que solo se fijan valores de los parámetros, el diagrama de la arquitectura, los patrones de entrada-salida y las ecuaciones diferenciales no difieren de las enunciadas para la Alternativa 2.1.

Resultados Numéricos: Realismo y Flexibilidad

Los resultados numéricos obtenidos para determinar el Realismo de esta arquitectura mostrados en el Cuadro 3.18 son inferiores a los obtenidos para el modelo Alternativa 2.1 (Cuadro 3.12) para los constrains 1 y 3, en los cuales se modela la presencia del estímulo de Fe³⁺. Esto, sumado al hecho de que no se presentan mejoras para el constraint 2, indica que la hipótesis introducida en esta arquitectura no permite explicar en mejor manera la activación del gen pbgP en presencia de los estímulos de Fe³⁺ y Mg²⁺. Consecuentemente con esta evidencia, no se observan soluciones que permitan determinar que esta es una arquitectura Flexible, como se muestra en el Cuadro 3.19.

Constraint	#Parámetros	Mg	Fe	mgta	pmrd	pbgp	Proporción	Probabilidad
Alternativa 2.4								
1	60	1	1	1	0	1	3,7403E-05	0,843752429
2	60	1	0	1	1	1	0,00014068	0,862588492
3	60	0	1	0	0	1	8,0215E-05	0,854550205
4	60	0	0	0	0	0	0,66870856	0,993315658

Cuadro 3.18: Resultados para el modelo Alternativa 2.4.

Constraint: número de la condición a cumplir, de acuerdo a lo descrito en la sección de modelos descubiertos. **#Parámetros:** cantidad de parámetros presentes en las ecuaciones del modelo. **Mg, Fe:** nodos cuyo valor equivale al estímulo de entrada para la red, representándose esquemáticamente con 0 y 1 la ausencia y presencia de estímulo, respectivamente. **mgta, pbgp, pmrd:** nodos cuyo valor equivale a la salida de la red, representándose con 1 para mgta y pbgp la activación de los sistemas PhoP/PhoQ y PmrA/PmrB, respectivamente, y con 0 para pmrd la influencia del sistema PhoP/PhoQ en PmrA/PmrB; 0 representa no actividad en todos los casos. **Proporción:** proporción de conjuntos de parámetros que cumplen con el constraint requerido del total de conjuntos generados. **Probabilidad:** calculada en función de la proporción, probabilidad de que al seleccionar un parámetro al azar para la red este le permita cumplir exitosamente con el constraint.

3.3. RESULTADOS OBTENIDOS DE LA APLICACIÓN DE GENIE 65

Modelo	#Parámetros	Mg	Fe	mgta	pmd	pbgp	Proporción	Probabilidad
Alternativa 2.4	60	1100	1010	1100	0100	1110	0	0

Cuadro 3.19: Resultados para la conjunción de los constraints 1 & 2 & 3 & 4 para el modelo Alternativa 2.4.

Para cada especie, se muestran las cuatro condiciones requeridas. Obsérvese que para un modelo particular, las primeras posiciones del array de valores para cada especie se refieren al constraint 1 expuesto anteriormente, y así sucesivamente.

3.3.8. Iteración 8: Enriquecido

Los modelos expuestos en las iteraciones anteriores permiten reproducir, parcialmente, el comportamiento global de la red genética conformada por los sistemas PhoP/PhoQ y PmrA/PmrB. Sin embargo, son insuficientes para reunir todos los requisitos del modelo para ser considerados como arquitecturas válidas para la red genética. Se requieren por lo tanto hipótesis adicionales.

Estas se refieren a dos puntos fundamentales para la interacción de ambos sistemas. El primero se relaciona con la definida conexión de ida, es decir el mecanismo mediante el cual la activación del sistema PhoP/PhoQ interactúa con el regulador de respuesta (PmrA) del sistema PmrA/PmrB para activar la transcripción de genes como pbgP. En base al fracaso de los modelos anteriores puede realizarse el siguiente razonamiento: la activación de pbgP en bajo Mg²⁺ (con estímulo) y bajo Fe³⁺ (sin estímulo) no puede explicarse exclusivamente en base a un cierto nivel basal de PhoP-P o a una diferencia entre las fuerzas de activación entre el monómero PmrA-P y el dímero PmrA-P₂, por lo que resulta coherente plantear la posibilidad de que ambas causas interactúen.

El segundo es relativo a la forma en la cual PmrB, en presencia de alto Fe³⁺ (el caso de PhoQ en presencia de bajo Mg²⁺ es análogo), aumenta la concentración de PmrA-P (PmrA fosforilado), activando la transcripción de genes tales como pbgP. Como se vio en el párrafo anterior, es necesario que exista la posibilidad de que PmrA se fosforile basalmente para explicar la activación del gen pbgP. Por lo tanto, la especie PmrB (favorecida en ausencia de Fe³⁺), debe poseer la capacidad de fosforilar a PmrA. Como se sabe que PmrB posee también el poder de defosforilar a PmrA-P, resulta coherente agregar esto al modelo. Ahora bien, la especie PmrB-ACT (favorecida en presencia de Fe³⁺) también debe poseer la capacidad de fosforilar (y consecuentemente, defosforilar) a PmrA. Se agregan entonces estos cuatro ejes a la arquitectura. Como son gobernados por distintas ecuaciones (con distintos parámetros), al explorar los conjuntos de parámetros posibles para la arquitectura es posible encontrar soluciones que favorezcan una u otra reacción, reproduciendo las diferencias observadas in vivo (se asigna un valor 0, es decir que se elimina el eje, para la defosforilación de PmrA-P mediada por PmrB-ACT dado que la evidencia experimental sugiere que esta reacción es despreciable comparada con la fosforilación mediada por PmrB). En este caso, la evidencia acumulada en las iteraciones transforma una hipótesis más restrictiva para la fosforilación, como la utilizada en el modelo Reducido, en la más general posible que se incorpora al modelo Enriquecido.

Al analizar lo expuesto en para las iteraciones anteriores, adicionando la información obtenida, pueden agregarse las siguientes reglas para el sistema (salvo mención explícita, no se diferencia entre especies fosforiladas y no

3.3. RESULTADOS OBTENIDOS DE LA APLICACIÓN DE GENIE 67

fosforiladas salvo en el punto 8):

1. El sistema PmrA/PmrB reprime la transcripción del gen pmrD.
2. La proteína PmrA se liga al promotor de pmrD.
3. El ligamiento de las proteínas PhoP y PmrA al promotor de pmrD no es mutuamente excluyente.
4. El sitio de ligamiento para PmrA en el promotor de pmrD es necesario para la represión de pmrD mediada por PmrA.
5. La proteína PhoP es reclutada en la región reguladora de los genes activados por PhoP en presencia de bajo Mg²⁺ pero no en presencia de alto Mg²⁺.
6. La eliminación de la autoregulación transcripcional positiva del sistema PhoP/PhoQ impide el ligamiento de la proteína PhoP a los promotores de un subconjunto de los genes regulados por la misma (pero no a todos).
7. Las proteínas PhoP y PmrA se ligan a sus promotores objetivo (por ejemplo, mgtA y pbgP, respectivamente) en respuesta al estímulo de bajo Mg²⁺.
8. El ligamiento a los promotores objetivo requiere de la forma fosforilada de PhoP, que es producida en presencia de bajo Mg²⁺.
9. La autoregulación positiva del operón phoP/phoQ amplifica la señal de bajo Mg²⁺.
10. La proteína PmrD se asocia a la forma fosforilada de la proteína PmrA, previniendo su defosforilación espontánea y la promovida por la proteína PmrB.
11. La afinidad de la forma fosforilada de PmrA por sus promotores objetivo es mayor que la de la forma no fosforilada.
12. La activación de la transcripción de los genes activados por PmrA depende de los mismos promotores tanto en el caso de alto Fe³⁺ como único estímulo (independiente del efecto de PmrD) como en el caso de bajo Mg²⁺ como único estímulo (dependiente del efecto de PmrD).
13. La proteína PmrD promueve el estado fosforilado de la proteína PmrA, no requiriéndose para este efecto otros componentes adicionales a la proteína PmrB.
14. La proteína PmrD inhibe la defosforilación de PmrA fosforilada promovida por PmrB.

68 *CAPÍTULO 3. APLICACIÓN DE GENIE A UNA RED GENÉTICA*

15. La proteína PmrD se liga a la PmrA fosforilada inhibiendo su defosforilación.
16. La proteína PmrB continuamente fosforila y defosforila a la proteína PmrA en condiciones de bajo Fe^{3+} , perdiendo la capacidad de defosforilarla en condiciones de alto Fe^{3+} .

Descripción de la Arquitectura

En base a lo expuesto anteriormente, se añaden al modelo las hipótesis generadas. La arquitectura queda entonces definida como puede observarse en el Fig.3.8.

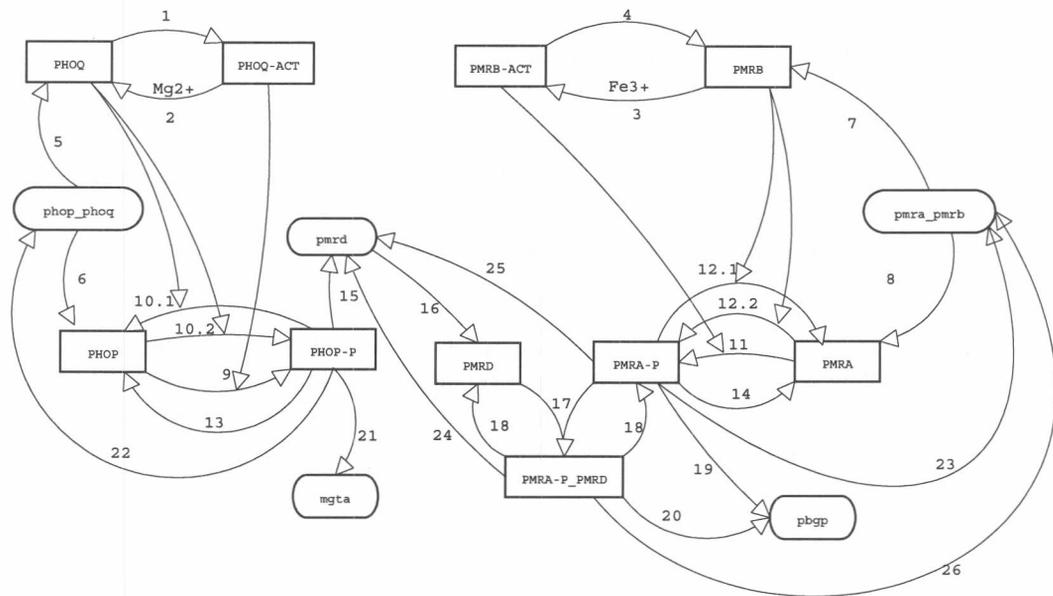


Figura 3.8: Modelo Enriquecido.

- 1 Bajo nivel de Mg²⁺ favorece el estado PHOQ-ACT (IVADO) en el equilibrio
- 2 Alto nivel de Mg²⁺ favorece el estado PHOQ en el equilibrio
- 3 Alto nivel de Fe³⁺ favorece el estado PMRB-ACT (IVADO) en el equilibrio
- 4 Bajo nivel de Fe³⁺ favorece el estado PMRB en el equilibrio
- 5 phop_phoq es traducido a PHOQ
- 6 phop_phoq es traducido a PHOP
- 7 pmra_pmr es traducido a PMRB
- 8 pmra_pmr es traducido a PMRA
- 9 PHOP es fosforilado a PHOP-P via PHOQ-ACT actividad kinasa
- 10.1 PHOP-P es defosforilado a PHOP via PHOQ actividad fosfatasa
- 10.2 PHOP es fosforilado a PHOP-P via PHOQ actividad kinasa
- 11 PMRA es fosforilado a PMRA-P via PMRB-ACT actividad kinasa
- 12.1 PMRA-P es defosforilado a PMRA via PMRB actividad fosfatasa
- 12.2 PMRA es fosforilado a PMRA-P via PMRB actividad kinasa
- 13 PHOP-P es defosforilado a PHOP via defosforilación espontánea
- 14 PMRA-P es defosforilado a PMRA via defosforilación espontánea
- 15 PHOP-P activa la transcripción de pmrd
- 16 pmrd es traducido a PMRD
- 17 PMRD se asocia con PMRA-P para constituir la especie PMRA-P_PMRD que activa la expresión de pbgp e inhibe la de pmrd pero no es afectada por la actividad fosfatasa de PMRB-ACT
- 18 PMRA-P_PMRD se disocia para formar PMRD y PMRA-P
- 19 PMRA-P activa la transcripción de pbgp
- 20 PMRA-P_PMRD activa la transcripción de pbgp
- 21 PHOP-P activa la transcripción de mgta
- 22 PHOP-P activa la transcripción de phop_phoq
- 23 PMRA-P activa la transcripción de pmra_pmr
- 24 PMRA-P_PMRD inhibe la transcripción de pmrd
- 25 PMRA-P inhibe la transcripción de pmrd
- 26 PMRA-P_PMRD activa la transcripción de phop_phoq

70 *CAPÍTULO 3. APLICACIÓN DE GENIE A UNA RED GENÉTICA*

Los patrones de entrada-salida para el modelo quedan reformulados como se muestra en el Cuadro 3.20.

Constraint	Mg	Fe	mgta	pmrd	pbgp
1	1	1	1	0	1
2	1	0	1	1	1
3	0	1	0	0	1
4	0	0	0	0	0

Cuadro 3.20: Patrones de entrada-salida para el modelo Enriquecido

Ecuaciones diferenciales para el modelo Enriquecido:

Utilizando el método descrito en el capítulo anterior, se genera el sistema de ecuaciones diferenciales para el modelo Enriquecido. Los superíndices en los términos de las ecuaciones indican la reacción del Cuadro 3.21 y de la Figura 3.8 de la cual se derivan.

3.3. RESULTADOS OBTENIDOS DE LA APLICACIÓN DE GENIE 71

Especie	Adición	Sustracción
phop_phoq	22	
PHOQ	2, 5	1
PHOQ-ACT	1	2
PHOP	6, 10.1,13	9, 10.2
PHOP-P	9, 10.2	10.1, 13
mgta	21	
pmra_pmr	23,26	
PMRB	7, 4	3
PMRB-ACT	3	4
PMRA	8, 12.1, 14	11
PMRA-P	11, 12.2,18	12.1, 14, 17
pbgp	19, 20	
PMRD	16, 18	17
PMRA-P_PMRD	17	18
pmrd	15	24,25

Cuadro 3.21: Términos de las ecuaciones que determinan las concentraciones de las especies.

Adición: Reacciones que determinan un término sustracción en la ecuación diferencial de la especie.

Sustracción: Reacciones que determinan un término de sustracción en la ecuación diferencial de la especie..

Decaimiento: A todas la especies se les agrega un término sustractivo extra para modelar el decaimiento de primer orden.

72 CAPÍTULO 3. APLICACIÓN DE GENIE A UNA RED GENÉTICA

$$\begin{aligned}
 \frac{d[\text{PHOQ}]}{dt} &= T_0 \left(\frac{1}{H_{\text{PHOQ}}} [\text{phop-phoq}]^{(5)} + r_{\text{PHOQACTmax}} \text{Mg2} [\text{PHOQACT}] [\text{Mg2}]^{(2)} \right. \\
 &\quad \left. - [\text{PHOQ}] r_{\text{PHOQ2PHOQACT}}^{(1)} - \frac{[\text{PHOQ}]^{(\text{dec})}}{H_{\text{PHOQ}}} \right) \\
 \frac{d[\text{phop-phoq}]}{dt} &= T_0 \left(\frac{1}{H_{\text{phop-phoq}}} \frac{[\text{PHOPP}]^{\nu_{\text{PHOPP-phop-phoq}}} + \text{PHOPP}^{\nu_{\text{PHOPP-phop-phoq}}}}{K_{\text{PHOPP-phop-phoq}}^{\nu_{\text{PHOPP-phop-phoq}}} + \text{PHOPP}^{\nu_{\text{PHOPP-phop-phoq}}}} \right)^{(22)} - \frac{[\text{phop-phoq}]^{(\text{dec})}}{H_{\text{phop-phoq}}} \\
 \frac{d[\text{PHOQ_ACT}]}{dt} &= T_0 \left([\text{PHOQ}] r_{\text{PHOQ2PHOQACT}}^{(1)} - r_{\text{PHOQACTmax}} \text{EMg2} [\text{PHOQACT}] [\text{EMg2}]^{(2)} \right. \\
 &\quad \left. - \frac{[\text{PHOQACT}]^{(\text{dec})}}{H_{\text{PHOQACT}}} \right) \\
 \frac{d[\text{PHOP}]}{dt} &= T_0 \left(\frac{1}{H_{\text{PHOP}}} [\text{phop-phoq}]^{(6)} \right. \\
 &\quad - P_{\text{PHOQACT}K} [\text{PHOP}] \frac{[\text{PHOQACT}]^{\nu_{\text{PHOQACT-PHOP}}}}{K_{\text{PHOQACT-PHOP}}^{\nu_{\text{PHOQACT-PHOP}}} + [\text{PHOQACT}]^{\nu_{\text{PHOQACT-PHOP}}}} \quad (9) \\
 &\quad + P_{\text{PHOQP}} [\text{PHOPP}] \frac{[\text{PHOQ}]^{\nu_{\text{PHOQ-PHOPP}}}}{K_{\text{PHOQ-PHOPP}}^{\nu_{\text{PHOQ-PHOPP}}} + [\text{PHOQ}]^{\nu_{\text{PHOQ-PHOPP}}}} \quad (10,1) \\
 &\quad - P_{\text{PHOQK}} [\text{PHOP}] \frac{[\text{PHOQ}]^{\nu_{\text{PHOQ-PHOP}}}}{K_{\text{PHOQ-PHOP}}^{\nu_{\text{PHOQ-PHOP}}} + [\text{PHOQ}]^{\nu_{\text{PHOQ-PHOP}}}} \quad (10,2) \\
 &\quad \left. + r_{\text{PHOPP}} [\text{PHOPP}]^{(13)} - \frac{[\text{PHOP}]^{(\text{dec})}}{H_{\text{PHOP}}} \right) \\
 \frac{d[\text{PHOPP}]}{dt} &= T_0 \left(P_{\text{PHOQACT}K} [\text{PHOP}] \frac{[\text{PHOQACT}]^{\nu_{\text{PHOQACT-PHOP}}}}{K_{\text{PHOQACT-PHOP}}^{\nu_{\text{PHOQACT-PHOP}}} + [\text{PHOQACT}]^{\nu_{\text{PHOQACT-PHOP}}}} \quad (9) \right. \\
 &\quad - P_{\text{PHOQP}} [\text{PHOPP}] \frac{[\text{PHOQ}]^{\nu_{\text{PHOQ-PHOPP}}}}{K_{\text{PHOQ-PHOPP}}^{\nu_{\text{PHOQ-PHOPP}}} + [\text{PHOQ}]^{\nu_{\text{PHOQ-PHOPP}}}} \quad (10,1) \\
 &\quad + P_{\text{PHOQK}} [\text{PHOP}] \frac{[\text{PHOQ}]^{\nu_{\text{PHOQ-PHOP}}}}{K_{\text{PHOQ-PHOP}}^{\nu_{\text{PHOQ-PHOP}}} + [\text{PHOQ}]^{\nu_{\text{PHOQ-PHOP}}}} \quad (10,2) \\
 &\quad \left. - r_{\text{PHOPP}} [\text{PHOPP}]^{(13)} - \frac{[\text{PHOPP}]^{(\text{dec})}}{H_{\text{PHOPP}}} \right) \\
 \frac{d[\text{mgta}]}{dt} &= T_0 \left(\frac{1}{H_{\text{mgta}}} \frac{[\text{PHOPP}]^{\nu_{\text{PHOPP-mgta}}} + [\text{PHOPP}]^{\nu_{\text{PHOPP-mgta}}}}{K_{\text{PHOPP-mgta}}^{\nu_{\text{PHOPP-mgta}}} + [\text{PHOPP}]^{\nu_{\text{PHOPP-mgta}}}} \right)^{(21)} - \frac{[\text{mgta}]^{(\text{dec})}}{H_{\text{mgta}}} \\
 \frac{d[\text{pmra-pmr}b]}{dt} &= T_0 \left(\frac{1}{H_{\text{pmra-pmr}b}} \left(1 - (1 - \alpha_{\text{PMRAP}}) \frac{1}{H_{\text{pmra-pmr}b}} \right. \right. \\
 &\quad \left. \frac{[\text{PMRAP}]^{\nu_{\text{PMRAP-pmr}a-pmr}b}}{K_{\text{PMRAP-pmr}a-pmr}b} + [\text{PMRAP}]^{\nu_{\text{PMRAP-pmr}a-pmr}b} \right) \\
 &\quad \left. (1 - \alpha_{\text{PMRAP-PMRD}}) \frac{1}{H_{\text{pmra-pmr}b}} \right. \\
 &\quad \left. \frac{[\text{PMRAP-PMRD}]^{\nu_{\text{PMRAP-PMRD-pmr}a-pmr}b}}{K_{\text{PMRAP-PMRD-pmr}a-pmr}b} + [\text{PMRAP-PMRD}]^{\nu_{\text{PMRAP-PMRD-pmr}a-pmr}b} \right) \quad (23,26) \\
 &\quad \left. - \frac{[\text{pmra-pmr}b]^{(\text{dec})}}{H_{\text{pmra-pmr}b}} \right) \\
 \frac{d[\text{PMRB}]}{dt} &= T_0 \left(\frac{1}{H_{\text{PMRB}}} [\text{pmrd}]^{(7)} + r_{\text{PMRBACT}} [\text{PMRBACT}]^{(4)} \right. \\
 &\quad \left. - r_{\text{PMRBACTmax}} \text{Fe3} [\text{PMRBACT}] [\text{Fe3+}]^{(3)} - \frac{[\text{PMRB}]^{(\text{dec})}}{H_{\text{PMRB}}} \right) \\
 \frac{d[\text{PMRBACT}]}{dt} &= T_0 \left(r_{\text{PMRBACTmax}} \text{Fe3} + [\text{PMRB}] [\text{Fe3+}]^{(3)} - [\text{PMRBACT}] r_{\text{PMRBACT}}^{(4)} \right. \\
 &\quad \left. - \frac{[\text{PMRBACT}]^{(\text{dec})}}{H_{\text{PMRB}}} \right) \\
 \frac{d[\text{PMRA}]}{dt} &= T_0 \left(\frac{1}{H_{\text{PMRA}}} [\text{pmra-pmr}b]^{(8)} + r_{\text{PMRAP}} [\text{PMRAP}]^{(14)} \right)
 \end{aligned}$$

3.3. RESULTADOS OBTENIDOS DE LA APLICACIÓN DE GENIE 73

$$\begin{aligned}
 & -P_{PMRBACT}K[PMRA] \frac{[PMRBACT]^{\nu} PMRBACT_PMRAP}{K_{PMRBACT_PMRAP}^{\nu} PMRBACT_PMRAP + [PMRBACT]^{\nu} PMRBACT_PMRAP} \quad (11) \\
 & + P_{PMRBK}[PMRAP] \frac{[PMRB]^{\nu} PMRB_PMRAP}{K_{PMRB_PMRAP}^{\nu} PMRB_PMRAP + [PMRB]^{\nu} PMRB_PMRAP} \quad (12,1) \\
 & - P_{PMRBK}[PMRA] \frac{[PMRB]^{\nu} PMRB_PMRA}{K_{PMRB_PMRA}^{\nu} PMRB_PMRA + [PMRB]^{\nu} PMRB_PMRA} \quad (12,2) \\
 & - \frac{[PMRA]^{(dec)}}{H_{PMRA}} \\
 \frac{d[PMRAP]}{dt} = & T_0(P_{PMRBACT}P[PMRA] \frac{[PMRBACT]^{\nu} PMRBACT_PMRA}{K_{PMRBACT_PMRA}^{\nu} PMRBACT_PMRA + [PMRBACT]^{\nu} PMRBACT_PMRA} \quad (11) \\
 & - P_{PMRBK}[PMRAP] \frac{[PMRB]^{\nu} PMRB_PMRAP}{K_{PMRB_PMRAP}^{\nu} PMRB_PMRAP + [PMRB]^{\nu} PMRB_PMRAP} \quad (12,1) \\
 & P_{PMRBK}[PMRA] \frac{[PMRB]^{\nu} PMRB_PMRA}{K_{PMRB_PMRA}^{\nu} PMRB_PMRA + [PMRB]^{\nu} PMRB_PMRA} \quad (12,2) \\
 & + \frac{[PMRAP_PMRD]^{(18)}}{H_{PMRAP_PMRD}} \\
 & - r_{PMRAP}[PMRAP]^{(14)} - r_{PMRAP_PMRD}max_PMRD[PMRAP][PMRD]^{(17)} \\
 & - \frac{[PMRAP]^{(dec)}}{H_{PMRAP}} \\
 \frac{d[pbgp]}{dt} = & T_0 \left(\frac{1}{H_{pbgp}} (1 - (1 - \alpha_{PMRAP} \frac{1}{H_{pbgp}} \frac{[PMRAP]^{\nu} PMRAP_pbgp}{K_{PMRAP_pbgp}^{\nu} PMRAP_pbgp + [PMRAP]^{\nu} PMRAP_pbgp} \right. \\
 & \left. (1 - \alpha_{PMRAP_PMRD} \frac{1}{H_{pbgp}} \frac{[PMRAP_PMRD]^{\nu} PMRAP_PMRD_pbgp}{K_{PMRAP_PMRD_pbgp}^{\nu} PMRAP_PMRD_pbgp + [PMRAP_PMRD]^{\nu} PMRAP_PMRD_pbgp} \right)^{(19,20)} \\
 & - \frac{[pbgp]^{(dec)}}{H_{pbgp}} \\
 \frac{d[PMRD]}{dt} = & T_0 \left(\frac{1}{H_{PMRD}} [pmrd]^{(16)} - r_{PMRAP_PMRD}max_PMRD[PMRAP][PMRD]^{(17)} \right. \\
 & \left. - \frac{[PMRD]^{(dec)}}{H_{PMRD}} \right) \\
 \frac{d[PMRAP_PMRD]}{dt} = & T_0 (r_{PMRAP_PMRD}max_PMRD[PMRAP][PMRD]^{(17)} \\
 & - \frac{[PMRAP_PMRD]^{(18)}}{H_{PMRAP_PMRD}}) \\
 \frac{d[pmrd]}{dt} = & T_0 \left(\frac{1}{H_{pmrd}} \frac{[PHOPP]^{\nu} PHOPP_pmrd}{K_{PHOPP_pmrd}^{\nu} PHOPP_pmrd + [PHOPP]^{\nu} PHOPP_pmrd} \right. \\
 & \left. (1 - \frac{[PMRAP]^{\nu} PMRAP_pmrd}{K_{PMRAP_pmrd}^{\nu} PMRAP_pmrd + [PMRAP]^{\nu} PMRAP_pmrd}) \right. \\
 & \left. (1 - \frac{[PMRAP_PMRD]^{\nu} PMRAP_pmrd}{K_{PMRAP_pmrd}^{\nu} PMRAP_pmrd + [PMRAP_PMRD]^{\nu} PMRAP_pmrd}) \right)^{(15,24,25)} - \frac{[pmrd]^{(dec)}}{H_{pmrd}}
 \end{aligned}$$

Resultados Numéricos: Realismo y Flexibilidad

Al generar una arquitectura en base a las nuevas hipótesis introducidas para el modelo Enriquecido, las proporciones de soluciones obtenidas para los constraints 1 a 3 aumentaron significativamente, como se observa en el Cuadro 3.22. Como puede observarse, el ranking de los constraints en base a las proporciones obtenidas, 4,2,3,1, es el mismo que el obtenido originalmente para el modelo Reducido (c.f. Cuadro 3.4), modificando el constraint 2 de manera tal que se requiera un crosstalk entre los dos subsistemas. Resulta significativo el hecho de que el modelo Enriquecido es el que presenta la menor proporción para el constraint 4, sugiriendo la existencia de un equilibrio entre la capacidad de la arquitectura para reproducir un cierto patrón en presencia de estímulos y su capacidad para no reproducir ningún patrón ante la ausencia de estos.

Inicialmente, como se aprecia en la primera línea del Cuadro 3.23, no se obtuvieron soluciones que permitieran concluir que esta es una arquitectura flexible. Esta falta de flexibilidad motivó un nuevo análisis de las condiciones a cumplir requeridas al modelo. Esta reformulación consiste en no requerir ningún valor especial para pmrD en los constraints 1 y 4. Esto reviste un especial interés porque se refiere a las condiciones para la interacción entre los dos subsistemas. En el caso del constraint 1, dado que se ejercen tanto el efecto positivo de PhoP-P como el negativo de PmrA-P sobre pmrD, no tiene sentido pedir que este sea ni especialmente "alto" ni especialmente "bajo". En el constraint 4, requerir que pmrD se "apague" en Mg=0 (alto Mg2+) impediría la posibilidad de que una cierta cantidad de PmrD constitutivo sea necesaria para el funcionamiento del sistema PmrA/PmrB. Como se observa en la segunda entrada del Cuadro 3.23, esta modificación derivó en la detección de soluciones para la conjunción de los cuatro constraints, permitiendo concluir que la arquitectura cumple con las condiciones de Flexibilidad propuestas.

Constraint	#Parámetros	Mg	Fe	mgta	pmrd	pbgp	Proporción	Probabilidad
1	68	1	1	1	0	1	0,000186349	0,88135701
2	68	1	0	1	1	1	0,001092771	0,90458411
3	68	0	1	0	0	1	0,000735495	0,89933251
4	68	0	0	0	0	0	0,469708768	0,98894913

Cuadro 3.22: Resultados para el modelo Enriquecido.

Constraint: número de la condición a cumplir, de acuerdo a lo descrito en la sección de modelos descubiertos. **#Parámetros:** cantidad de parámetros presentes en las ecuaciones del modelo. **Mg, Fe:** nodos cuyo valor equivale al estímulo de entrada para la red, representándose esquemáticamente con 0 y 1 la ausencia y presencia de estímulo, respectivamente. **mgta, pbgp, pmrd:** nodos cuyo valor equivale a la salida de la red, representándose con 1 para mgta y pbgp la activación de los sistemas PhoP/PhoQ y PmrA/PmrB, respectivamente, y con 1 para pmrd la influencia del sistema PhoP/PhoQ en PmrA/PmrB; 0 representa no actividad en todos los casos. **Proporción:** proporción de conjuntos de parámetros que cumplen con el constraint requerido del total de conjuntos generados. **Probabilidad:** calculada en función de la proporción, probabilidad de que al seleccionar un parámetro al azar para la red este le permita cumplir exitosamente con el constraint.

3.3. RESULTADOS OBTENIDOS DE LA APLICACIÓN DE GENIE 75

Modelo	#Parámetros	Mg	Fe	mgta	pmrd	pbgp	Proporción	Probabilidad
Alternativa 3	68	1 1 0 0	1 0 1 0	1 1 0 0	0 1 0 0	1 1 1 0	0	0
Alternativa 3*	68	1 1 0 0	1 0 1 0	1 1 0 0	- 1 0 -	1 1 1 0	4,88464E-06	0,835400248

Cuadro 3.23: Resultados para la conjunción de los constraints 1 & 2 & 3 & 4 para el modelo Enriquecido.

Para cada especie, se muestran las cuatro condiciones requeridas. Obsérvese que para un modelo particular, las primeras posiciones del array de valores para cada especie se refieren al constraint 1 expuesto anteriormente, y así sucesivamente.
 *Se eliminan dos condiciones referidas a pmrd.

Resultados Numéricos: Robustez

En la Fig.3.9 se grafica la puntuación obtenida por la red al modificar un parámetro particular de una solución (y dejando los demás fijos), haciendo que este tome valores a lo largo de todo el intervalo definido como biológicamente significativo. Para las pruebas se tomó una solución escogida aleatoriamente entre las que reproducen la conjunción de los constraints 1 a 4. Se considera que la red reproduce el patrón requerido si obtiene un score no mayor a 0.3. Puede apreciarse entonces que algunos parámetros pueden tomar valores con varios órdenes de magnitud de diferencia, evidenciando una robustez de la red con respecto a los mismos, como por ejemplo ν_{PHOQ_PHOP} . Otros parámetros muestran una región acotada fuera de la cual la red pierde el comportamiento deseado, por ejemplo H_{PHOP} , y finalmente existe un subconjunto en el cual pequeñas perturbaciones impactan en forma significativamente negativa en la red empobreciendo los scores obtenidos, como es el caso de K_{PHOP-P_pmrd} .

En el caso de los parámetros robustos de la red, resulta de interés el hecho de determinar si la robustez de un parámetro se debe o no a compensaciones en los valores de los demás, es decir valores inusualmente extremos de otro parámetro que disminuyan la importancia del parámetro robusto dentro de la solución. Tomemos como ejemplo a los parámetros $P_{PHOQACTK}$, que determina la fuerza de PhoQ-ACT para fosforilar a PhoP, y K_{PHOPP_mgtA} , que determina la fuerza con la cual la especie PhoP-P se asocia al DNA de *mgtA* activando su transcripción. Un valor inusualmente grande de $P_{PHOQACTK}$, que incrementaría en gran medida la concentración de PhoP-P, podría enmascarar la importancia de K_{PHOPP_mgtA} haciendo que siempre se activara la transcripción de *mgtA*, con una consecuente robustez de la solución con respecto a este último parámetro. Las pruebas realizadas muestran que no existen correlaciones significativas de a pares entre los valores de los parámetros en las distintas soluciones, por lo que de existir esta compensación se debería a combinaciones de más de dos parámetros.

3.3. RESULTADOS OBTENIDOS DE LA APLICACIÓN DE GENIE 77

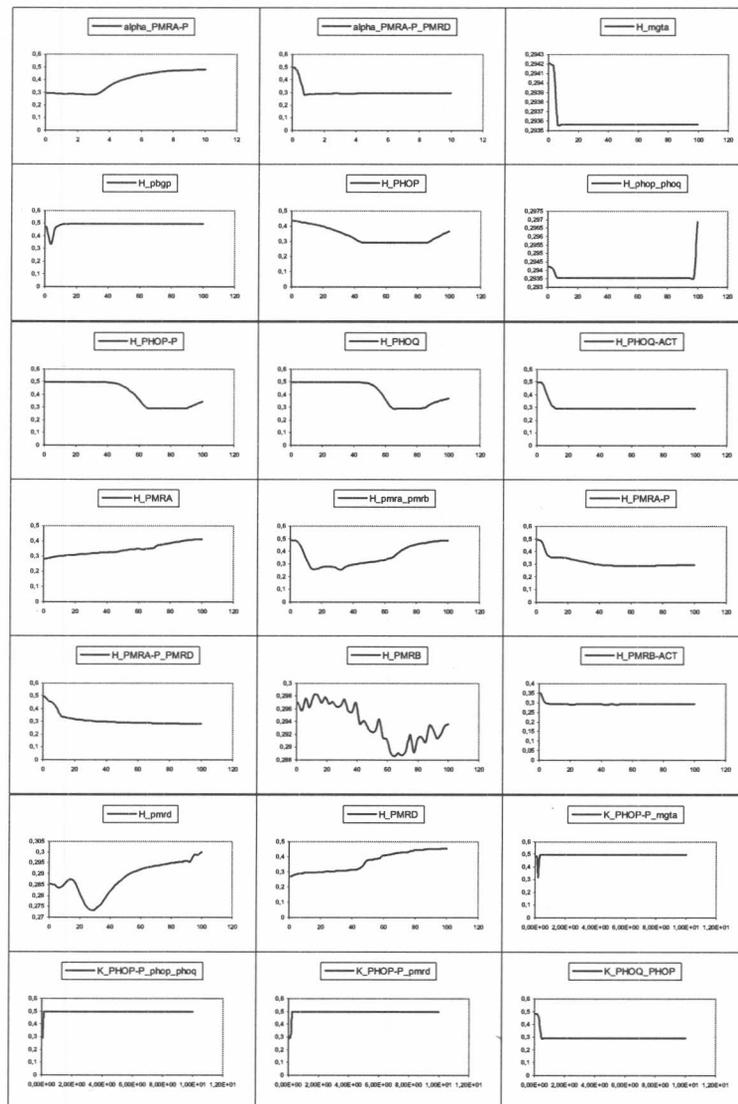


Figura 3.9: Robustez de los parámetros.

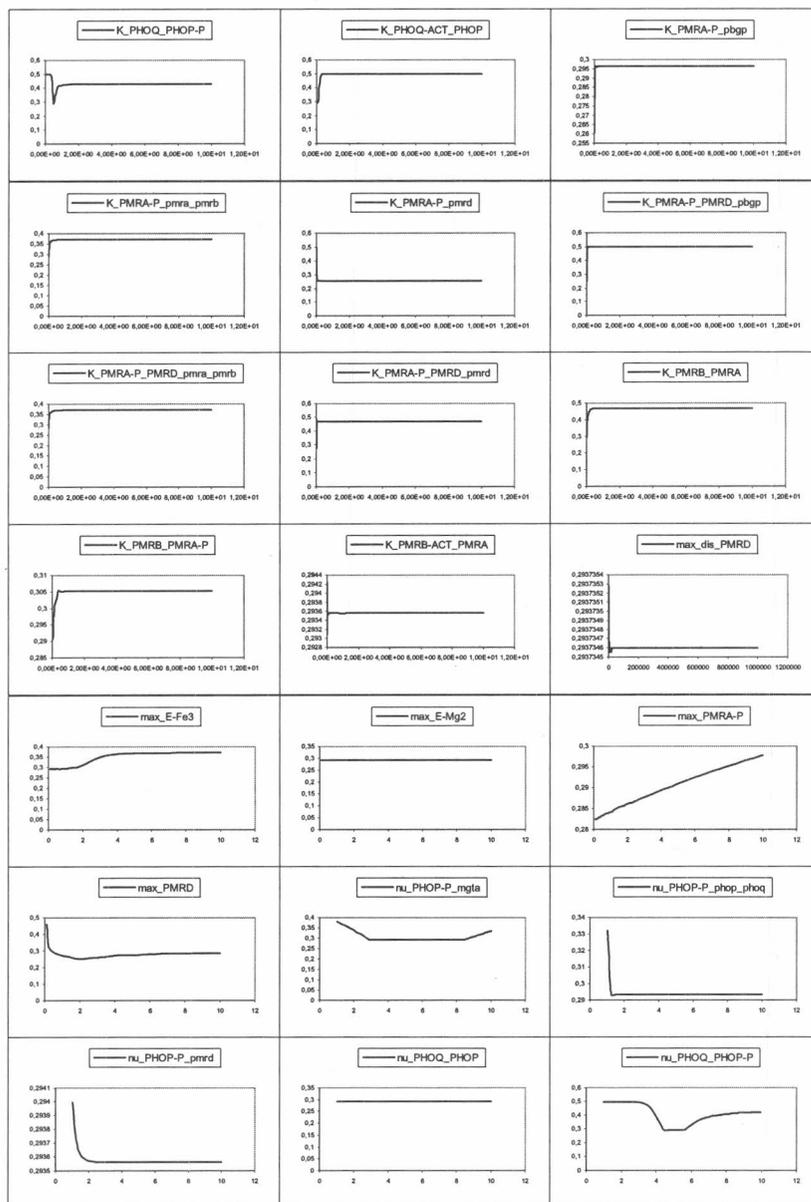


Figura 3.10: Robustez de los parámetros (continuación).

3.3. RESULTADOS OBTENIDOS DE LA APLICACIÓN DE GENIE 79

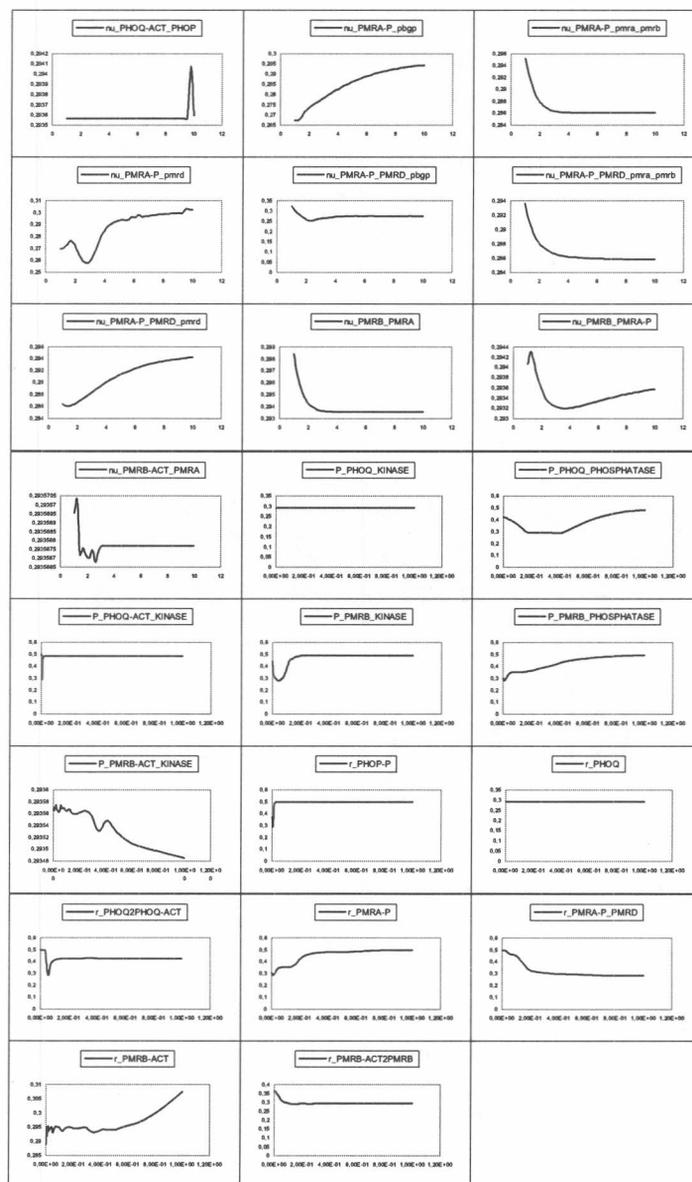


Figura 3.11: Robustez de los parámetros (continuación).

En la Fig.3.12 se muestran los resultados de realizar el análisis de sensibilidad de los parámetros anteriormente descritos en base a diez soluciones tomadas al azar para cada restricción en el modelo Enriquecido. Los parámetros se encuentran agrupados por tipo: vidas medias (prefijo H_o), valor medio de activación máxima de la transcripción (prefijo K_o), fosforilación (prefijo P_o), dimerización (prefijo r_o) y coeficiente de Hill (prefijo nu_o). En el Cuadro 3.24 se listan aquellos parámetros puntuales que tienen un comportamiento “poco robusto” en estos experimentos. Las pruebas de la Fig.3.12 fueron repetidas tomando como parámetros poco robustos a aquellos que sólo pueden tomar valores en menos de un 5% del intervalo propuesto sin hacer que el score de la solución sobrepase un cierto umbral, aunque no se muestran los resultados debido a que la proporción resultó 0 para todos los constraints.

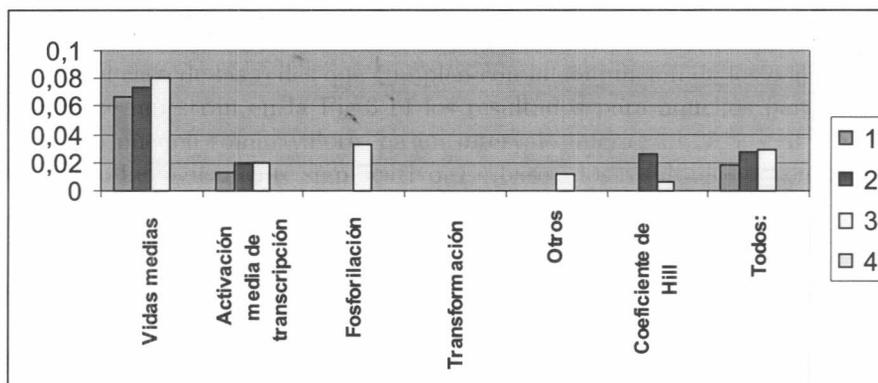


Figura 3.12: Proporción de cada tipo de parámetro que sólo puede tomar valores en un intervalo inferior al 25% del intervalo biológicamente significativo, por constraint para el modelo Enriquecido.

En la Fig.3.13 se realiza la misma prueba, pero se muestran aquellos parámetros que pueden tomar valores en el 100% del intervalo definido como biológicamente significativo sin que el conjunto de parámetros al que pertenecen pierda la capacidad de reproducir los patrones pedidos. En el Cuadro 3.24 se muestran las proporciones para los parámetros no robustos más significativos en la solución escogida.

En el caso de las redes que cumplen con la conjunción de los constraints 1...4, se muestran en la Fig.3.14 los resultados para aquellos parámetros que sólo pueden tomar valores en un intervalo inferior al 25% y al 5% del intervalo biológicamente significativo. En base a los resultados obtenidos, es posible concluir que la arquitectura del modelo Enriquecido es **robusta con respecto a sus ejes**.

3.3. RESULTADOS OBTENIDOS DE LA APLICACIÓN DE GENIE 81

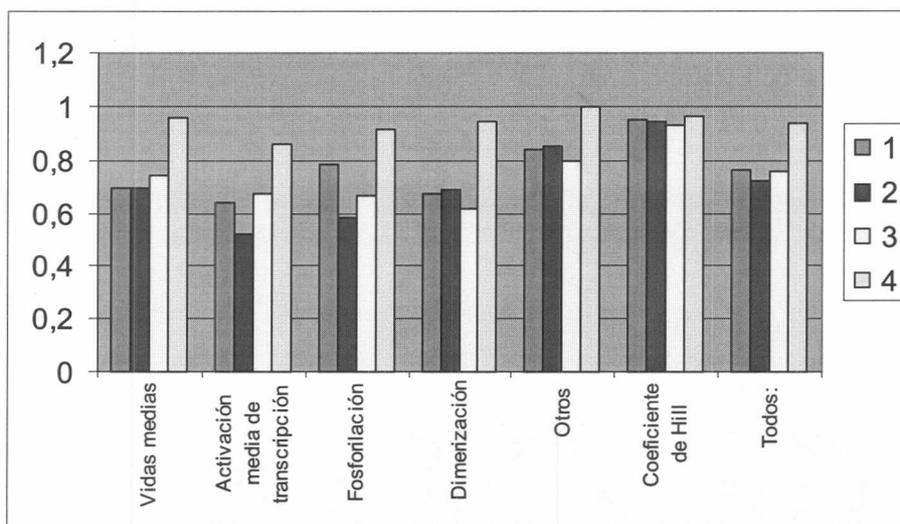


Figura 3.13: Proporción de cada tipo de parámetro que puede tomar valores en el 100 % del intervalo biológicamente significativo, por constraint para el modelo Enriquecido.

Constraint					
1		2		3	4
H_pbgp	1.0	H_pbgp	1.0	H_pbgp	1.0
K_PMRA-P_pbgp	0.1	K_PMRA-P_pbgp	0.2	H_pmra_pmrB	0.2
K_PMRB_PMRA	0.1	nu_PHOP-P_phop_phoq	0.2	K_PMRA-P_pbgp	0.2
		H_PHOQ	0.1	K_PMRB_PMRA	0.1
		K_PMRA-P_PMRD_pbgp	0.1	max_E-Mg2	0.1
		nu_PHOQ_PHOP	0.1	nu_PMRA-P_pmrD	0.1
		nu_PMRB_PMRA	0.1	P_PHOQ_PHOSPHATASE	0.1
				P_PMRB_KINASE	0.1

Cuadro 3.24: Proporción de cada parámetro particular que sólo puede tomar valores en un intervalo inferior al 25 % del intervalo biológicamente significativo.

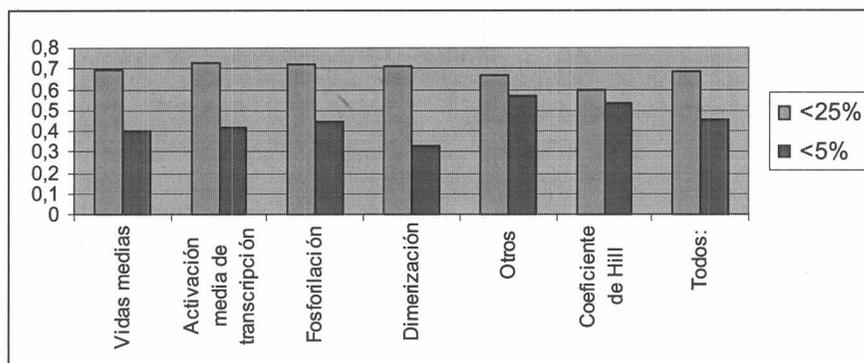


Figura 3.14: Proporción de cada tipo de parámetro que sólo puede tomar valores en un intervalo inferior al 25 % del intervalo biológicamente significativo para la conjunción de los constraints en el modelo Enriquecido.

3.3. RESULTADOS OBTENIDOS DE LA APLICACIÓN DE GENIE 83

En el Cuadro 3.25 se muestran los resultados de simulaciones sobre soluciones obtenidas al evaluar realismo y flexibilidad, a las cuales se les han introducido cambios en los valores iniciales de las especies (nodos de la red). La disminución de la proporción de soluciones para la arquitectura que se observa al someter a la arquitectura a valores **reducidos** de las concentraciones iniciales permiten inferir que el modelo Enriquecido es más sensible a bajas concentraciones iniciales de las especies que a altas concentraciones y a la presencia de ruido en los valores. Asimismo, puede verse que la restricción 4 es la menos sensible a variaciones en los valores iniciales de los nodos, mientras que las restricciones 1, 2 y 3 se comportan en forma relativamente similar ante estas variaciones.

Estos resultados permiten determinar que la arquitectura encontrada en la octava iteración, el denominado modelo Enriquecido, es robusta con respecto a los valores iniciales de los nodos de la red, es decir **robusta con respecto los nodos**.

Condición Inicial	Constraint				
	1	2	3	4	1&2&3&4
Mínimo	0,821	0,882	0,873	1	0,473
Fuerte	0,992	1	1	1	0,971
Degradado	0,973	0,992	0,983	1	0,975

Cuadro 3.25: Proporción de soluciones válidas al modificar los valores iniciales de las especies.

Proporción de soluciones para los distintos constraints del modelo Enriquecido que mantienen un score exitoso luego de modificar los valores iniciales de las concentraciones de las especies. **Mínimo**: los valores de las concentraciones de las especies se colocan en cero salvo los referentes a mrna, que presentan un nivel inicial bajo, y Fe y Mg donde 0.2 es considerado bajo y 0.8 alto (cumpliendo con el constraint correspondiente). **Fuerte**: todas las especies comienzan con un valor bajo salvo Fe y Mg, donde 0.0 es bajo y 1.0 alto coherentemente con el constraint particular. **Degradado**: se agrega a todas las especies ruido normal con media igual al diez por ciento del valor utilizado para generar la solución.

3.4. Resumen de los Resultados Numéricos

En esta sección se presenta un resumen de los resultados numéricos mostrados en las secciones anteriores. Esto permite obtener una visión global y comparativa de los resultados de la aplicación de GENIE.

3.4.1. Realismo

En el Cuadro 3.26 los resultados de las pruebas de Realismo para todas las arquitecturas obtenidas. En este puede apreciarse que todas las arquitecturas son realistas, en el sentido de que se observan soluciones válidas para todos los constraints, aunque en distinta medida. En todos las arquitecturas, salvo la Alternativa 2.1 por escaso margen, puede verse que el constraint número 1 es el que resulta más difícil de cumplir, en el sentido de que se observan las menores proporciones de soluciones. En todos los casos el constraint 4 es el más sencillo de cumplir por las arquitecturas.

Modelo	Constraint	#Parámetros	Mg	Fe	mgta	pmrd	pbgp	Proporción	Probabilidad
Reducido	1	60	1	1	1	0	1	0,00035474	0,875988521
	2	60	1	0	1	1	0	0,06442205	0,955323753
	3	60	0	1	0	0	1	0,0019469	0,901202504
	4	60	0	0	0	0	0	0,66514196	0,993227127
Reducido Sin Feedback	1	56	1	1	1	1	1	0,00097532	0,883557379
	2	56	1	0	1	1	0	0,12128993	0,963029804
	3	56	0	1	0	0	1	0,00321704	0,902589529
	4	56	0	0	0	0	0	0,66439315	0,992725141
Alternativa 1	1	60	1	1	1	0	1	0,0008419	0,8886981
	2	60	1	0	1	1	1	0,0001273	0,8611505
	3	60	0	1	0	0	1	0,0036405	0,9106524
	4	60	0	0	0	0	0	0,6597033	0,9930912
Alternativa 2.1	1	62	1	1	1	0	1	0,00010728	0,86293171
	2	62	1	0	1	1	1	0,00010331	0,862405877
	3	62	0	1	0	0	1	0,00017975	0,870144576
	4	62	0	0	0	0	0	0,79079849	0,996221478
Alternativa 2.2	1	58	1	1	1	0	1	8,2774E-05	0,850391367
	2	58	1	0	1	1	1	0,00012101	0,855977336
	3	58	0	1	0	0	1	0,00046723	0,876149225
	4	58	0	0	0	0	0	0,78856455	0,995912841
Alternativa 2.3	1	60	1	1	1	0	1	4,8346E-05	0,847369295
	2	60	1	0	1	1	1	5,0393E-05	0,84795508
	3	60	0	1	0	0	1	0,0005726	0,883006975
	4	60	0	0	0	0	0	0,30381418	0,980340175
Alternativa 2.4	1	60	1	1	1	0	1	3,7403E-05	0,843752429
	2	60	1	0	1	1	1	0,00014068	0,862588492
	3	60	0	1	0	0	1	8,0215E-05	0,854550205
	4	60	0	0	0	0	0	0,66870856	0,993315658
Enriquecido	1	68	1	1	1	0	1	0,000186349	0,88135701
	2	68	1	0	1	1	1	0,001092771	0,90458411
	3	68	0	1	0	0	1	0,000735495	0,89933251
	4	68	0	0	0	0	0	0,469708768	0,98894913

Cuadro 3.26: Resultados del los tests de realismo para todos los modelos.

Constraint: número de la condición a cumplir, de acuerdo a lo descrito en la sección de modelos descubiertos. **#Parámetros:** cantidad de parámetros presentes en las ecuaciones del modelo. **Mg, Fe:** nodos cuyo valor equivale al estímulo de entrada para la red, representándose esquemáticamente con 0 y 1 la ausencia y presencia de estímulo, respectivamente. **mgta, pbgp, pmrd:** nodos cuyo valor equivale a la salida de la red, representándose con 1 para mgta y pbgp la activación de los sistemas PhoP/PhoQ y PmrA/PmrB, respectivamente, y con 0 para pmrd la influencia del sistema PhoP/PhoQ en PmrA/PmrB; 0 representa no actividad en todos los casos. **Proporción:** proporción de conjuntos de parámetros que cumplen con el constraint requerido del total de conjuntos generados. **Probabilidad:** calculada en función de la proporción, probabilidad de que al seleccionar un parámetro al azar para la red este le permita cumplir exitosamente con el constraint.

En la Figura 3.15 se muestra una comparación de las proporciones obtenidas para los modelos Reducido y Reducido sin Feedback. Estas resultan similares para todos los constraints salvo el número 2, relativo a la activación del sistema PhoP/PhoQ e inactivación del sistema PmrA/PmrB en condiciones de bajo Mg²⁺ y bajo Fe³⁺. Esto resulta coherente con el hecho de que la anulación del feedback negativo, es decir la habilidad de la proteína PmrA-P de inhibir la activación del gen pmrD, facilita la activación de este último gen, lo cual es una condición necesaria dentro del constraint. En la Figura 3.16 se muestran las probabilidades asociadas a las frecuencias obtenidas.

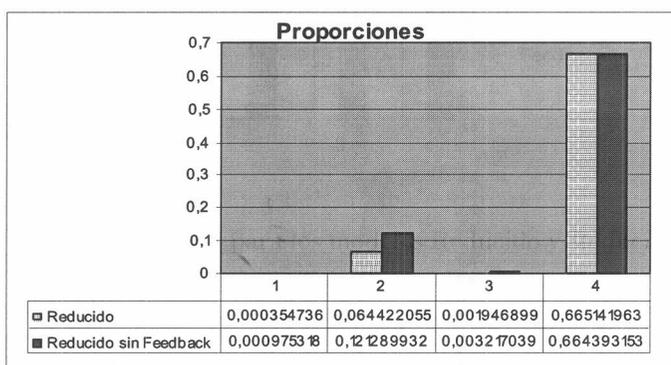


Figura 3.15: Proporciones para los modelos Reducido y Reducido sin Feedback.

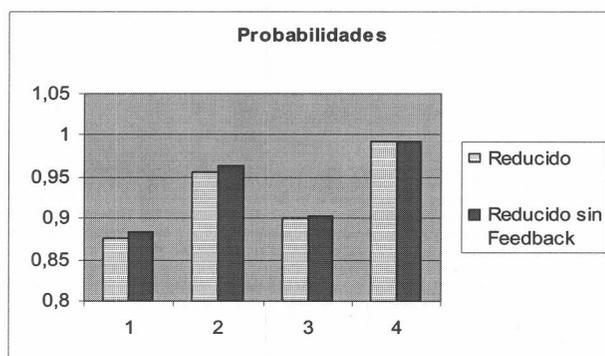


Figura 3.16: Probabilidades para los modelos Reducido y Reducido sin Feedback.

Las arquitecturas Alternativa 2.1, 2.2, 2.3 y 2.4 (Figura 3.17) presentan distintas proporciones para los constraints requeridos. En el caso del constraint 1, las arquitecturas 2.1 y 2.2 presentan proporciones similares que doblan aproximadamente a las obtenidas para las arquitecturas 2.3 y 2.4, evidenciando que la anulación del feedback negativo no impacta significativamente en este constraint, pero sí lo hacen la anulación del feedback positivo de PhoP-P y la utilización de una proporción fija para diferenciar la influencia de la proteína PmrA-P y el dímero PmrA-P_PmrD en la activación del gen *pbgP*. En el constraint 2 se observa que la anulación del feedback positivo (arquitectura Alternativa 2.3) impacta negativamente en la habilidad de la red para cumplirlo. Este hecho es coherente con la noción de que el feedback positivo es necesario para potenciar el efecto del Mg^{2+} , activando el sistema PmrA/PmrB. El constraint 3, por el contrario, es más fácilmente reproducido por la red al anular el feedback positivo. En la Figura 3.18 se muestran las probabilidades asociadas.

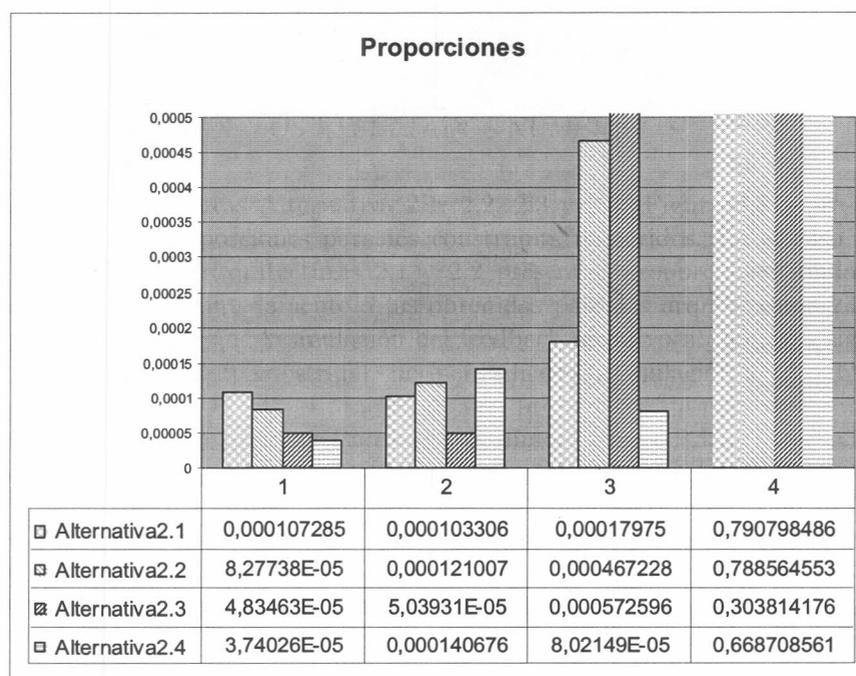


Figura 3.17: Proporciones para los modelos Alternativa 2.

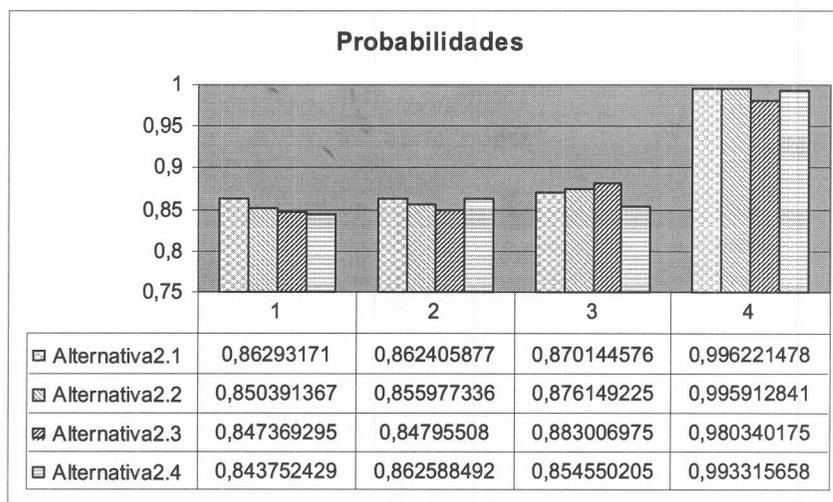


Figura 3.18: Probabilidades para los modelos Alternativa 2.

En las Figuras 3.19 y 3.20 se muestran las proporciones y probabilidades obtenidas para el modelo Enriquecido.

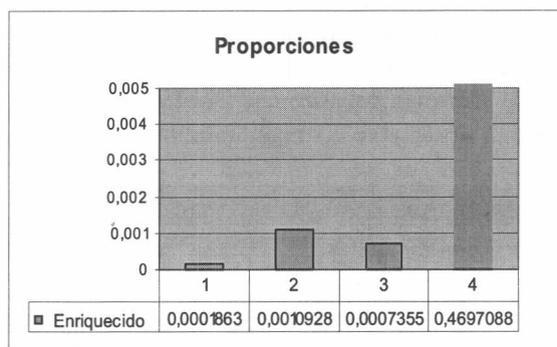


Figura 3.19: Proporciones para el modelo Enriquecido.

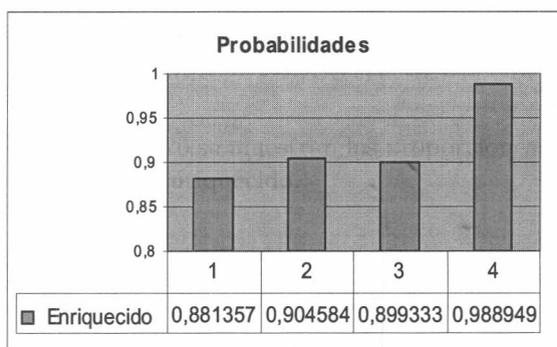


Figura 3.20: Probabilidades para el modelo Enriquecido.

3.4.2. Flexibilidad

En el Cuadro 3.27 y en la Figura 3.21 puede verse claramente que sólo las arquitecturas de los modelos Reducido, Reducido sin Feedback y Enriquecido cumplen con lo requerimientos definidos para considerarlas Flexibles, aunque sólomente la arquitectura del modelo Enriquecido es flexible incluyendo un crosstalk entre los subsistemas PhoP/PhoQ y PmrA/PmrB. En la Figura 3.22 pueden apreciarse las probabilidades asociadas.

Modelo	#Parámetros	Mg	Fe	mgta	pmrd	pbgp	Proporción	Probabilidad
Reducido	60	1 1 0 0	1 0 1 0	1 1 0 0	0 1 0 0	1 0 1 0	1,87764E-05	0,834116778
Reducido sin Feedback	56	1 1 0 0	1 0 1 0	1 1 0 0	1 1 0 0	1 0 1 0	3,97356E-05	0,834476142
Reducido sin Feedback*	56	1 1 0 0	1 0 1 0	1 1 0 0	0 1 0 0	1 0 1 0	0	0
Alternativa 1	60	1 1 0 0	1 0 1 0	1 1 0 0	0 1 0 0	1 1 1 0	0	0
Alternativa 2.1	62	1 1 0 0	1 0 1 0	1 1 0 0	0 1 0 0	1 1 1 0	0	0
Alternativa 2.2	58	1 1 0 0	1 0 1 0	1 1 0 0	0 1 0 0	1 1 1 0	0	0
Alternativa 2.3	60	1 1 0 0	1 0 1 0	1 1 0 0	0 1 0 0	1 1 1 0	0	0
Alternativa 2.4	60	1 1 0 0	1 0 1 0	1 1 0 0	0 1 0 0	1 1 1 0	0	0
Alternativa 3	68	1 1 0 0	1 0 1 0	1 1 0 0	0 1 0 0	1 1 1 0	0	0
Alternativa 3**	68	1 1 0 0	1 0 1 0	1 1 0 0	- 1 0 -	1 1 1 0	4,88464E-06	0,835400248

Cuadro 3.27: Resultados para la conjunción de los constraints 1 & 2 & 3 & 4.

Para cada especie, se muestran las cuatro condiciones requeridas. Obsérvese que para un modelo particular, las primeras posiciones del array de valores para cada especie se refieren al constraint 1 expuesto anteriormente, y así sucesivamente.

*Conjunción de constraints iguales a los requeridos para el modelo Reducido.

**Se eliminan dos condiciones referidas a pmrd.

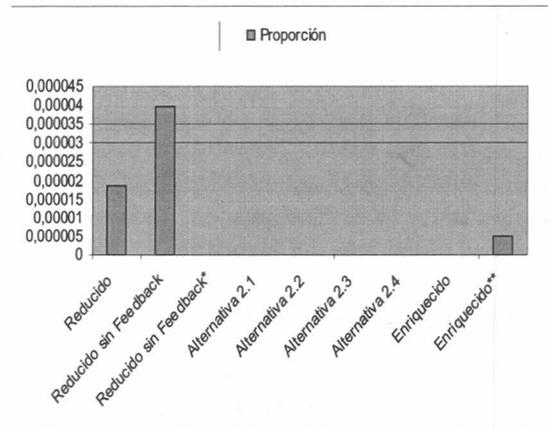


Figura 3.21: Proporciones para la conjunción de los constraints.

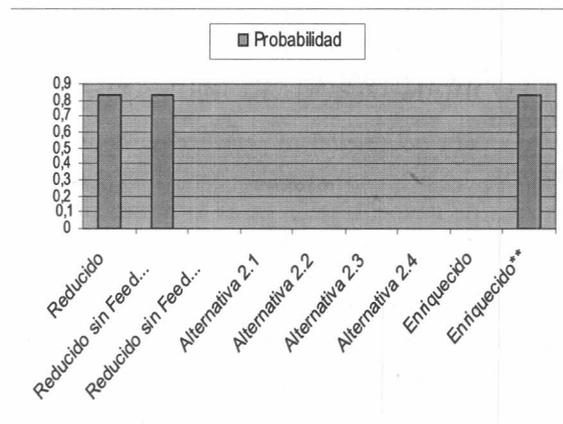


Figura 3.22: Probabilidades para la conjunción de los constraints.

3.5. Predicciones y Validación Experimental Biológica

Los conjuntos de parámetros obtenidos para las distintas arquitecturas propuestas permiten, en distinta medida, reproducir patrones genéricos determinados de antemano como biológicamente significativos. Se utilizaron los resultados y predicciones del modelo Enriquecido, dado que este superó con éxito las distintas pruebas numéricas del algoritmo.

3.5.1. Binding: Experimentos ChIP

Con el fin de determinar con mayor precisión si los resultados obtenidos en las simulaciones se ajustan a lo observado experimentalmente, se realizaron ensayos de Chromatin Immunoprecipitation (ChIP), que muestran el nivel de ocupación relativa de phoP en los sitios de transcripción de los genes phoP, mgtA y pmrD (Cuadro 3.28). Estos experimentos permiten determinar la cantidad de activador, PhoP en este caso, que se asocia al DNA de cada gen regulado, phoP, mgtA y pmrD, cuantificando la fuerza de la asociación. La transcripción de los genes activados a RNA y la consecuente traducción a proteína puede considerarse como una función de la fuerza de asociación del activador.

Asumiendo que esta ocupación se correlaciona con un aumento de la transcripción [DZG03], se determinaron los coeficientes de correlación de Pearson entre las curvas de concentración en el experimento de ChIP y las obtenidas en las simulaciones (Cuadro 3.29).

Tiempo(min)	Valores de inmuno precipitación relativa		
	phoP	mgtA	pmrD
0	0,08	0,213	0
5	0,177	0,273	0,077
10	0,329	0,455	0,137
20	0,639	0,889	0,375
30	0,589	0,933	0,411
60	0,461	1,035	0,374
90	0,139	0,515	0,152

Cuadro 3.28: Ocupación relativa de phoP en los sitios de transcripción de phoP, mgtA y pmrD en condiciones de bajo magnesio y bajo hierro (constraint 2).

Al comparar los Cuadros 3.28 y 3.29, puede observarse que las especies medidas experimentalmente sufren un decaimiento en sus concentraciones a partir de un cierto tiempo de medición. En el caso del RNA del gen mgtA, esta disminución no puede explicarse sólo en base a la ecuación

Constraint	Coeficiente de correlación de Pearson					
	Hasta los 60 min**			Hasta los 90 min**		
	mgtA	pmrD	phoP	mgtA	pmrD	phoP
2*	0,862	0,472	0,832	0,474	0,12	0,44
AND* (constraint 2)	0,937	0,738	0,709	0,789	0,376	0,196

Cuadro 3.29: Correlación entre las curvas obtenidas en las simulaciones y las obtenidas en experimentos de inmunoprecipitación que determinan la ocupación relativa de phoP en los sitios de transcripción de phoP, mgtA y pmrD.

* Con el fin de ejemplificar los resultados obtenidos, se eligió al azar una solución entre el 10% de las mejores soluciones obtenidas en el modelo Enriquecido. ** Se calcula el coeficiente de Pearson a los 60 y 90 minutos debido a que a partir del minuto 60 se observa una marcada disminución en la concentración de todas las especies estudiadas. Este hecho se observó al estudiar los resultados de las pruebas biológicas, posteriormente a la determinación de los patrones a aprender por el modelo, y por lo tanto no tiene sentido requerir que las soluciones obtenidas lo cumplan.

diferencial utilizada para modelarla (Sección 3.3.8), dado que esta no posee elementos que modelen una función que tenga un máximo temporal y luego disminuya hasta un cierto valor. Sin embargo, como en la arquitectura propuesta, y en su sistema de ecuaciones diferenciales asociado, se modela el comportamiento de toda la red, es posible proponer que esta disminución se debe a un decrecimiento en la concentración de la especie activadora de la transcripción del gen mgtA (PhoP-P, c.f Fig.3.8). Ahora bien, como en la arquitectura la especie mgtA modela genéricamente a un gen regulado positivamente por PhoP-P, y dado que la evidencia de binding indica que existen otros genes activados por PhoP, puede realizarse la siguiente pregunta: es posible que el modelo obtenido prediga el comportamiento, es decir la curva de expresión temporal, de estos genes activados por PhoP al modificar los parámetros que determinan la vida media de PhoP-P y la expresión de nuestro gen genérico, mgtA? En la actualidad, es conocido que estos genes son activados por PhoP, aunque no está determinado se trata de la especie fosforilada o defosforilada. Una conclusión basada en los resultados obtenidos, que será enunciada en el capítulo siguiente, es que la regulación debe basarse en la especie fosforilada, por lo que aquí se utiliza a PhoP fosforilado (PhoP-P) como especie activadora. Estas predicciones pueden observarse en el Cuadro 3.30 y la Figura 3.23.

Parámetros Modificados	1	2	3	4	5	6	7	8
H_PHOP-P	10	10	10	10	10	10	10	10
H_mgta	20	20	20	20	20	20	20	20
nu_mgta	1	1	1	1	1	5	5	5
K_mgta	0,01	0,025	0,05	0,075	0,1	0,01	0,025	0,05
Tiempo Simulación	1	2	3	4	5	6	7	8
1,2849354	0,05896046	0,05465868	0,04873491	0,04397124	0,04005691	0,06222641	0,06222318	0,06212229
5,5535517	0,2265214	0,2062486	0,17956446	0,15905434	0,1427833	0,24245845	0,24240884	0,24086916
9,701928	0,35406262	0,31687745	0,27002367	0,23546538	0,20886762	0,38435948	0,38410547	0,37643117
20,884008	0,57254815	0,48889568	0,39483467	0,3320416	0,2869061	0,64797634	0,64316016	0,538118
30,610004	0,66758	0,5489413	0,42610356	0,34948128	0,29677925	0,78332984	0,76105815	0,4793043
61,0401	0,73003554	0,54423565	0,384785	0,2986733	0,2444688	0,9487208	0,7032462	0,1681631
90,17029	0,68447155	0,4714436	0,312123	0,23386288	0,23717758	0,964926	0,47416095	0,04873141
Parámetros Modificados	9	10	11	12	13	14	15	
H_PHOP-P	10	10	10	10	10	10	10	
H_mgta	20	20	20	20	20	20	20	
nu_mgta	5	5	10	10	10	10	10	
K_mgta	0,075	0,1	0,01	0,025	0,05	0,075	0,1	
Tiempo	9	10	11	12	13	14	15	
1,2849354	0,06144471	0,0590678	0,06222644	0,06222644	0,06222625	0,06221568	0,06203599	
5,5535517	0,231076	0,20256801	0,24245897	0,24245894	0,24244386	0,24159466	0,2288557	
9,701928	0,3351071	0,25425678	0,38436213	0,38436183	0,38405693	0,36872455	0,27224776	
20,884008	0,3325371	0,19487628	0,6480268	0,64795446	0,5918878	0,31448683	0,16777533	
30,610004	0,2349884	0,12759757	0,7835725	0,78229743	0,4781341	0,19665623	0,10335163	
61,0401	0,06066377	0,03011525	0,95271015	0,789251	0,11450987	0,04313001	0,02258061	
90,17029	0,00154114	0,00732108	0,98811007	0,29802316	0,02684142	0,01005408	0,00526253	

Cuadro 3.30: Curvas de expresión predichas por GENIE para distintos genes regulados por PhoP al modificar los parámetros de la solución.

Para realizar las predicciones, se seleccionó de entre las soluciones válidas para el modelo Enriquecido la de menor vida media de PHOP-P (H_PHOP-P). Luego se generaron quince combinaciones de los parámetros de activación media maximal (K_mgta) y coeficiente de cooperación de Hill (nu_mgta) para el gen genérico mgta. En la tabla se muestran, para cada simulación, los valores de expresión obtenidos a distintos tiempos.

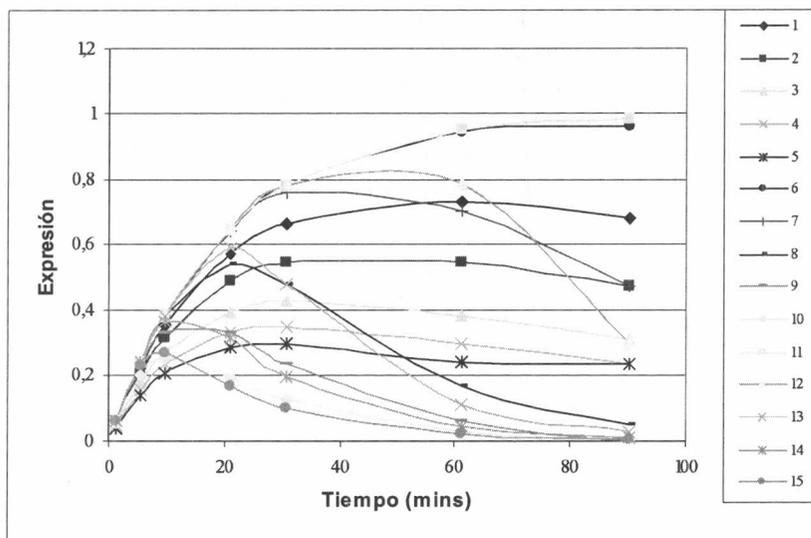


Figura 3.23: Curvas de expresión obtenidas para genes regulados por PhoP al modificar los valores de los parámetros de su ecuación diferencial asociada.

Con el fin de validar experimentalmente las predicciones del modelo, se realizaron distintos experimentos de Chromatin Inmuno Precipitation (ChIP), siendo sus resultados los expuestos en la Fig. 3.24, donde se incluyen los resultados del Cuadro 3.28. Puede observarse que los valores de binding graficados la Fig. 3.24 indican que los genes presentan distintos tiempos de aparición y desaparición de sus productos, así como también distintos valores máximos de expresión.

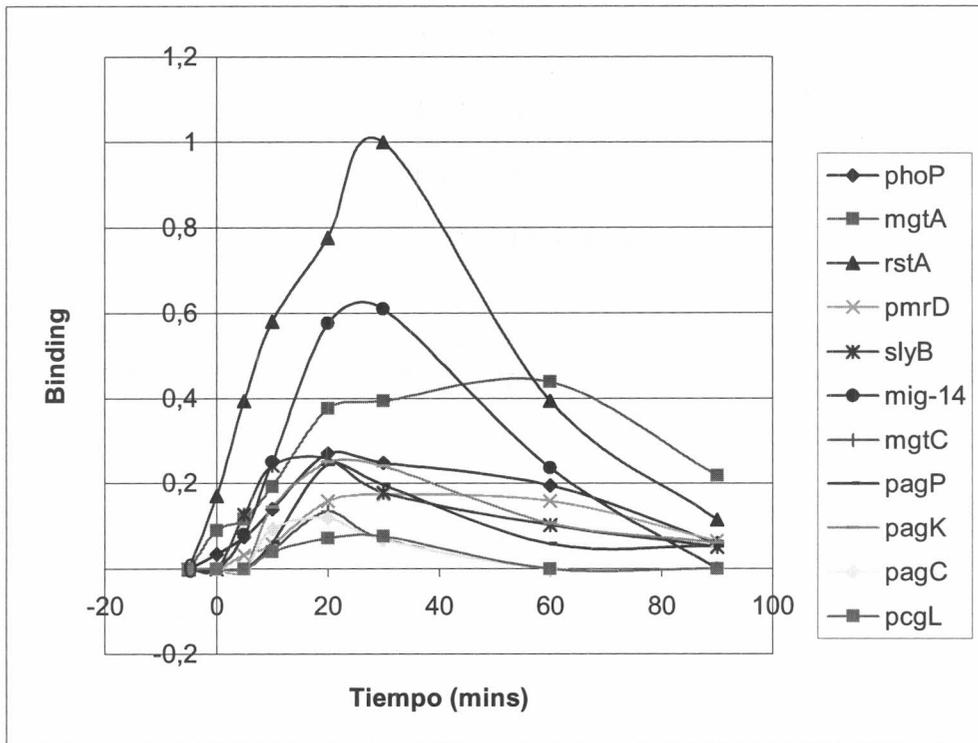


Figura 3.24: Binding de PhoP a la región promotora de distintos genes en función del tiempo.

Curvas de binding obtenidas para un experimento de CHIP analizando el binding relativo de PhoP (sin distinguir entre la especie fosforilada y la no fosforilada) a distintos genes. Los valores de binding fueron normalizados dividiéndolos por el mayor valor presente, para obtener valores entre 0 y 1 comparables a los predichos.

Posteriormente, se realizó un clustering de los resultados obtenidos para la ocupación temporal de los binding sites, cuyos resultados se observan en la Figura 3.25. En esta se observan al menos tres segmentos que implican un distinto orden en el cual los genes comenzarían a transcribirse. Esto puede asociarse a distintas funciones biológicas de los genes de estos tres grupos. Por ejemplo, los genes del grupo que se activa primero pueden estar asociados a mecanismos de respuesta rápida de la bacteria ante un cambio repentino en el entorno como el que se produce al encontrarse dentro de un macrófago, y los restantes a mecanismos que no requieran tal celeridad.

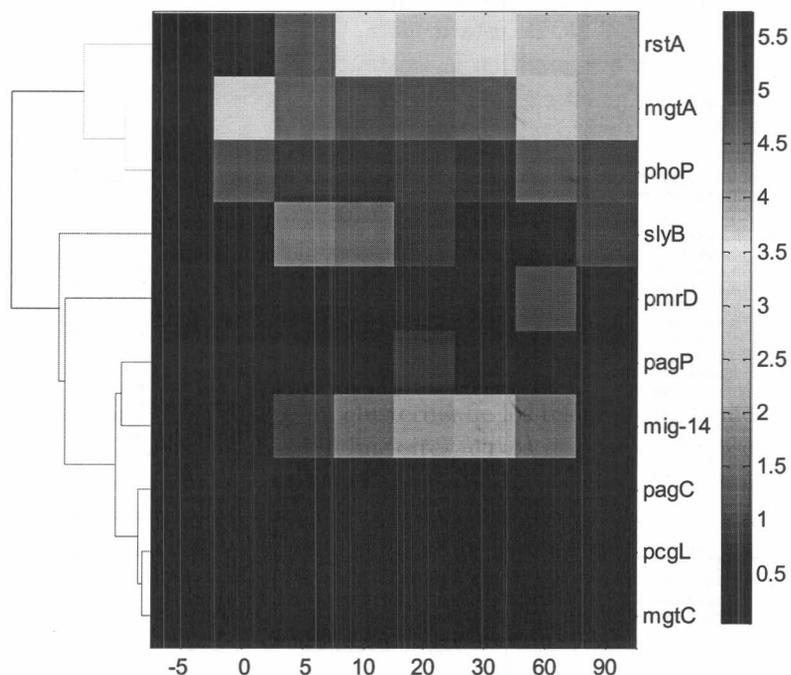


Figura 3.25: Clustering jerárquico sobre los datos de los experimentos de CHIP.

El dendrograma muestra el clustering jerárquico de genes y el orden temporal. La IP relativa se mide para diez genes como una función de ocho tiempos configurando una matriz de entrada de 10x8 puntos, la cual se normaliza por la media relativa de IP para cada tiempo. Luego, se aplicó una técnica de clustering jerárquico, combinando una variedad de medidas inter e intra cluster, en las cuales la Correlación de Pearson y conectividad completa, respectivamente, proveen la mejor combinación estimada por el coeficiente de correlación copenhética.

En el Cuadro 3.31 puede verse un análisis de la correlación entre las curvas predichas por los modelos surgidos de la modificación de los parámetros de la ecuación de expresión de *mgtA* (modelando la expresión de nuevos genes hipotéticos) y las obtenidas experimentalmente. Para todas las curvas experimentales se obtiene una alta correlación, lo cual certifica la validez de las predicciones. Más aún, se observa que los distintos genes pueden agruparse de acuerdo a su correlación con una misma curva, siendo posible agrupar las curvas de expresión temporales obtenidas experimentalmente para los genes regulados por PhoP en base a su similaridad con tres de las curvas predichas. Esta agrupación es consistente con la obtenida al realizar el clustering de la Fig. 3.25. De esta manera, GENIE permite formular modelos para el sistema que predicen la existencia de tres patrones generales de expresión para los genes regulados por uno de los elementos de la red, la proteína PhoP-P.

	phoP	mgtA	rstA	pnrD	slyB	mig-14
1	0,61928709	0,85794978	0,33981874	0,82900028	0,26492335	0,48246059
2	0,74576937	0,90652632	0,50200732	0,90152895	0,42424263	0,62833589
3	0,84385231	0,91513845	0,64837599	0,93632341	0,58115667	0,75165981
4	0,88615311	0,89874129	0,72579435	0,93636486	0,67187368	0,8116634
5	0,81397157	0,84270697	0,65716106	0,88230393	0,64110106	0,74696069
6	0,48457839	0,78407633	0,18265797	0,73523995	0,11935587	0,33487934
7	0,85897756	0,96137959	0,6485592	0,9710322	0,49942263	0,75927577
8	0,84352487	0,49976237	0,94611432	0,62670515	0,92220393	0,92202348
9	0,55070375	0,11792405	0,74954543	0,24777596	0,9178142	0,65077755
10	0,28659218	-0,1507399	0,54579546	-0,0273447	0,79958289	0,39458262
11	0,46897421	0,77224203	0,1670765	0,72251153	0,10739606	0,31967558
12	0,9130103	0,97569855	0,71356089	0,98141336	0,53376224	0,80470675
13	0,80105453	0,42983635	0,91562486	0,56325458	0,91377424	0,89447241
14	0,42572242	-0,0079116	0,63967864	0,11352873	0,88694265	0,52465998
15	0,12272233	-0,2913141	0,40322016	-0,1797741	0,69813201	0,22577177
Max Val	0,91301	0,975699	0,946114	0,981413	0,922204	0,922023
Simulación	12	12	8	12	8	8
2 ^o Mejor	4	7	13	7	9	13
Simulación						
	mgtC	pagP	pagK	pagC	pcgL	
1	0,27716907	0,55674403	0,5991426	0,18018035	0,32097919	
2	0,74576937	0,67109489	0,72241444	0,33232432	0,47099036	
3	0,55281042	0,7628738	0,82205318	0,48287298	0,60867876	
4	0,62534764	0,8035574	0,86710168	0,56966587	0,68207881	
5	0,60067196	0,78255531	0,83482741	0,54888883	0,64997028	
6	0,14282156	0,4369099	0,47037	0,04220185	0,17710809	
7	0,50921551	0,74791195	0,80498073	0,41359311	0,58302558	
8	0,89572365	0,835517	0,8686341	0,90023069	0,9324809	
9	0,73294236	0,53028742	0,59107194	0,85041257	0,73332059	
10	0,51280449	0,25338073	0,32540817	0,68762304	0,50931899	
11	0,13228195	0,42638344	0,45821103	0,03199406	0,16522582	
12	0,50909639	0,71588885	0,79386547	0,42468576	0,58693351	
13	0,91940832	0,83067834	0,84277191	0,91707216	0,93035611	
14	0,64766201	0,4095504	0,4853395	0,81048726	0,63651178	
15	0,34921301	0,07199035	0,15531988	0,55652528	0,34925742	
Max Val	0,919408	0,835517	0,868634	0,917072	0,932481	
Simulación	13	8	8	13	8	
2 ^o Mejor	8	13	4	8	13	
Simulación						

Cuadro 3.31: Correlación entre las curvas predichas por GENIE y las curvas de binding obtenidas experimentalmente para distintos genes regulados por PhoP.

Coefficiente de correlación de Pearson calculado para cada par predicción-gen. Se indica el número de la simulación que presenta la mayor correlación con la curva de binding obtenida experimentalmente. Se muestra también el número de la simulación que posee el segundo mayor valor de correlación. Nótese que las curvas de expresión se asocian a tres simulaciones (8, 12 y 13), de manera similar a la observada realizando un clustering sobre los datos experimentales.

3.5.2. Expresión: Experimentos de Real-Time PCR

En la sección anterior, se realizó una validación experimental de las predicciones del modelo, asumiendo que un mayor tiempo de ocupación de la región reguladora de un cierto gen por el activador implicaba una mayor transcripción. Con el fin de validar las predicciones con valores reales de expresión, se realizaron experimentos de Real Time PCR, observándose las curvas obtenidas en la Fig.3.26. Estos experimentos permiten medir en distintos momentos la cantidad de mRNA relativo a cada uno de los genes estudiados, cuantificando la expresión de los mismos.

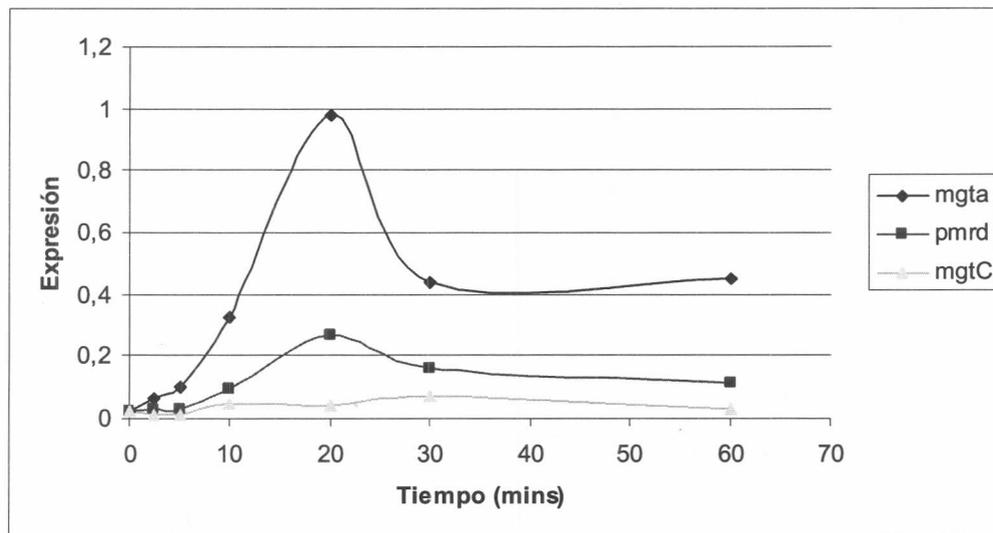


Figura 3.26: Resultados de expresión para los genes mgtA, pmrD y mgtC medidos mediante Real-Time PCR.

Curvas de binding obtenidas para un experimento de Real Time PCR analizando la expresión de mrna de distintos genes. Los valores de expresión fueron normalizados dividiéndolos por el mayor valor presente, para obtener valores entre 0 y 1 comparables a los predichos.

En este caso, también se observa una alta correlación entre las curvas predichas y las obtenidas experimentalmente (Cuadro 3.32). Nuevamente, puede observarse un comportamiento similar, es una alta correlación con una misma curva predicha, para los genes mgtA y pmrD, no observada en el experimento de clustering.

	mgta	pmrd	mgcC
1	0,7428926	0,75430198	0,67429949
2	0,78215364	0,80021453	0,72122241
3	0,81301839	0,83651635	0,75740664
4	0,81301839	0,83651635	0,75740664
5	0,83210404	0,85965176	0,77972152
6	0,69802534	0,70199739	0,61993537
7	0,77609027	0,80462668	0,7492911
8	0,79808435	0,85600374	0,78825444
9	0,54665623	0,57044854	0,49490447
10	0,22711315	0,22772401	0,19115841
11	0,69670407	0,70032166	0,61787471
12	0,74874118	0,77005989	0,71043908
13	0,78781614	0,8444247	0,74230568
14	0,43118399	0,43980575	0,37572926
15	0,03416355	0,02025974	0,01934497
Max Val	0,832104	0,859652	0,788254
Simulación	5	5	8

Cuadro 3.32: Correlación entre las curvas predichas y las curvas de expresión obtenidas experimentalmente.

Coefficiente de correlación de Pearson calculado para cada par predicción-gen. Se indica el número de la simulación que presenta la mayor similitud con la curva de expresión obtenida experimentalmente.

3.5.3. Materiales y Métodos

Se realizó un ensayo de serie temporal ChIP con la cepa EG13922 criada en un medio N-mínimo a pH 7.7 con 10 mM Mg^{2+} y luego cambiada a un medio con 50 μM Mg^{2+} . Se tomaron alícuotas inmediatamente antes y después del cambio al medio de bajo Mg^{2+} . Se realizó un ensayo de PCR promotor-específica tanto en el DNA inmunoprecipitado (IP) como en el total (input), cuantificado utilizando phosphoimager. Un conjunto adicional de PCRs fue realizado utilizando cada conjunto de primers y DNA total serialmente diluido para determinar que la señal de IP se encontraba en rango lineal. Los valores de IP relativa fueron obtenidos dividiendo las señales inmunoprecipitadas dentro del rango lineal de PCR dividiendo por la correspondiente señal de input.

3.6. Observaciones Finales

La aplicación de GENIE a un problema concreto de la biología molecular muestra el poder del método para modelar redes de expresión genética, no sólo al permitir extraer conclusiones sobre la dinámica del sistema y sus propiedades estructurales, sino también al formalizar los pasos que permiten especificar el problema en términos computacionales contrastables, generando y confirmando o refutando nuevas hipótesis sobre el mismo.

Las técnicas de análisis de binding sites utilizadas permiten obtener un primer conjunto especies pertenecientes a la red, y de interacciones entre las mismas. Estas reglas son luego enriquecidas en base a la bibliografía existente, permitiendo capturar evidencia sobre interacciones entre las especies a nivel post-transcripcional. Finalmente, el algoritmo completa los puntos no conocidos sobre el problema considerando la combinatoria de todas las reglas posibles en base a librerías de módulos funcionales.

Las reglas obtenidas son mapeadas mediante el algoritmo en un conjunto de arquitecturas posibles para la red, las cuales son testeadas iterativamente de acuerdo a distintos criterios, como se ha visto en las iteraciones enunciadas en este capítulo. Una vez que el algoritmo encuentra una arquitectura factible para el problema estudiado, esta es utilizada para determinar propiedades sistémicas emergentes del modelo, como su robustez con respecto a los parámetros y a los valores iniciales de las concentraciones de las especies. Asimismo, esta puede ser utilizada para realizar predicciones sobre el comportamiento dinámico de la red, mostrándose un ejemplo en el cual GENIE predice la existencia de grupos de genes regulados por especies de la red modelada (la proteína PhoP-P, en este caso). Como se ve en la Sección 3.5, estas predicciones son validadas experimentalmente, mostrándose el valor de GENIE como generador automático de nuevas hipótesis no consideradas hasta su aplicación.

Capítulo 4

Conclusiones

4.1. Introducción

Los experimentos realizados sobre los distintos modelos propuestos permiten extraer diversas conclusiones con respecto al sistema estudiado. Desde el punto de vista de la biología celular se muestra un modelo soportado experimentalmente (biológica y computacionalmente) para la interacción a nivel transcripcional y post-transcripcional de dos sistemas de dos componentes. Esta interacción permite a la célula combinar distintas funcionalidades asociadas a una misma función biológica en base a la interacción de distintos estímulos. De esta manera, se constituye una agregación que conforma una verdadera red genética, que por su simpleza relativa resulta óptima para el análisis.

Desde la óptica computacional, resulta de interés el proceso de formalización de un problema complejo derivado de la biología molecular, permitiendo transformar una serie de hipótesis biológicas (con distinto nivel de soporte experimental) en un sistema “testable” computacionalmente, en el sentido de formular hipótesis sobre su funcionamiento y luego verificarlas o no. Esta formalización permite la formulación y verificación de hipótesis a nuevos niveles, como las referidas a propiedades inherentes a la arquitectura del sistema.

4.2. GENIE Predice Interacciones que Permiten Explicar los Resultados Experimentales

Las arquitecturas propuestas permiten explicar el comportamiento observado *in vivo* para la interacción entre los sistemas de dos componentes PmrA/PmrB y PhoP/PhoQ [KG04]. Como puede observarse en el Cuadro 3.4, el modelo reducido cumple con los constraints 1 a 4 individualmente. Asimismo, como puede verse en el Cuadro 3.27, esta arquitectura permite obtener soluciones para la conjunción de los cuatro constraints. Sin embar-

go, el constraint 2 utilizado para la búsqueda de soluciones indica que se requiere que pbgP no se encuentre activado en presencia de $Mg=1$ (bajo Mg^{2+}) y $Fe=0$. En siguientes iteraciones del análisis, la reformulación de la evidencia existente indicó que el valor biológicamente esperado para pbgP en este caso era 1 (activado) y no 0. Al someter al modelo a este nuevo constraint, y al requerir el cumplimiento de la conjunción de los constraints 1 a 4, no se observaron soluciones válidas.

El cambio en el constraint 2 que se describe en el párrafo anterior fue introducido en el análisis de los restantes modelos, impactando en la proporción de soluciones obtenidas. Efectivamente, puede verse que para los resultados de los modelos Alternativa2.*, la proporción de soluciones para los constraints 1 a 3 cae un orden de magnitud con respecto al Reducido, salvo para el constraint 4 en el cual aumenta. Esto último resulta coherente dado que, si el constraint 2, que marca el efecto de la “conexión de ida” entre los subsistemas, es más difícil de aprender, es esperable un aumento en la proporción de soluciones para el constraint 4 que justamente indica la “no actividad” de los sistemas ante falta de estímulos (es decir, el “blanco” del experimento).

El modelo Reducido Sin Feedback (Cuadro 3.7) permite obtener soluciones para los constraints individuales 1 a 4 utilizados para el modelo Reducido, y también modificando el constraint 1 (llamémoslo 1') para que no se requiera la represión de pmrD ($RNA\ pmrD=0$) en el caso $Mg=1$ y $Fe=1$ (esto es consecuente con la anulación del feedback negativo). Si bien esto daría la impresión de que el modelo puede producir soluciones tanto para el constraint 1 como el 1' (lo cual parecería contradictorio), al ajustar los constraints requiriendo que el modelo produzca soluciones para las conjunciones 1&2&3&4 y 1'&2&3&4 (Cuadro 3.8), sólo se observaron soluciones válidas en el segundo caso. Al igual que con el modelo Enriquecido, al modificar el constraint 2 no se observaron soluciones válidas.

Dentro del conjunto de modelos Alternativa2.*, se exploran distintas modificaciones sobre el modelo Reducido. En el caso Alternativa2.1 se introduce la autoregulación positiva de phoP mediante PhoP-P (PhoP fosforilado). La eliminación de elementos a este modelo (la mencionada autoregulación positiva, o el feedback negativo anteriormente descrito) produce un impacto negativo sobre la frecuencia de soluciones encontradas (Cuadro 3.26), dado que estas disminuyen en varios casos en un orden de magnitud. Sin embargo, ninguno de los modelos Alternativa2.* permite encontrar soluciones para la conjunción de los constraints (Cuadro 3.27).

Al estudiar las proporciones de soluciones encontradas para los distintos constraints, puede observarse que estas no son uniformes. Si bien existen diferencias entre las distintas arquitecturas, resulta claro que el constraint 4 presenta una proporción de soluciones muy superior al resto, es decir que resulta más sencillo de aprender para la red. Esto tiene sentido dado que el constraint 4 representa la condición de que la red no active ninguno de

los nodos de salida (las especies RNA de *mgtA*, *pmrD* y *pbgP*) al no existir estímulos de los nodos de entrada (Mg^{2+} y Fe^{3+}), es decir que la red no presente actividad. La importancia de este requerimiento surge al someter a la red a pruebas que involucran el cumplimiento de los cuatro constraints simultáneamente, en las cuales el número 4 funciona como test “blanco”, asegurando que la actividad evidenciada en los otros tres no se produce trivialmente, independientemente de los estímulos provocados. El constraint 1, contrariamente al 4, presenta la menor proporción de soluciones para todas las arquitecturas propuestas. Esto resulta coherente al considerar que en este caso se requiere que, ante el estímulo combinado del hierro y el magnesio, se evidencie la activación (es decir, la transducción de las señales) en ambos subsistemas, aumentando la transcripción de *mgtA* y *pbgP* y disminuyendo la transcripción de *pmrD* debido al feedback negativo de PmrA-P. Los constraints 2 y 3 se ubican en una posición intermedia. De estos resultados puede apreciarse que cada patrón a seguir por la red plantea exigencias de distinta magnitud sobre interacción entre las especies que se traducen en distintos subconjuntos de valores para los parámetros de las ecuaciones diferenciales. Como se verá en la siguiente sección, estos subconjuntos no son mutuamente excluyentes, y existen soluciones para la red con la suficiente flexibilidad como para reproducir cualquiera de los patrones requeridos.

4.2.1. El Modelo Enriquecido Predice Exhaustivamente la Funcionalidad de la Red

Las distintas hipótesis alternativas generadas en base a los resultados anteriormente descriptos, conjuntamente con el reprocesamiento de la evidencia experimental disponible, permitieron generar el modelo que denominamos Enriquecido, que incluye elementos tales como una nueva hipótesis para la interacción de PmrD con PmrA-P y una nueva hipótesis para la acción fosfatasa - kinasa de las especies PmrB y PhoQ. Dentro de las distintas hipótesis alternativas formuladas en el transcurso de las iteraciones anteriores, se incluyeron las que permitían describir en mejor medida los resultados experimentales.

Como se observa en el Cuadro 3.22, el modelo enriquecido permite encontrar soluciones para los constraints 1 a 4. Sin embargo, como se ve en la anteúltima entrada del Cuadro 3.23, inicialmente no se encontraron soluciones para este modelo. Esto motivó un nuevo análisis de las condiciones a cumplir requeridas a los distintos modelos. De entre los propuestos, el único para el cual se encontraron soluciones luego de una reformulación resultó ser el Enriquecido.

La reformulación de los constraints consiste en no requerir ningún valor especial para *pmrD* en los constraints 1 y 4. Esto reviste un especial interés porque se refiere a las condiciones para la interacción entre los dos subsistemas. En el caso del constraint 1, dado que se ejercen tanto el efec-

to positivo de PhoP-P como el negativo de PmrA-P sobre pmrD, no tiene sentido pedir que este sea ni especialmente “alto” ni especialmente “bajo”. En el constraint 4, requerir que pmrD se “apague” en Mg=0 (alto Mg²⁺) impediría la posibilidad de que una cierta cantidad de PmrD constitutivo sea necesaria para el funcionamiento del sistema PmrA/PmrB.

4.3. PmrA/PmrB y PhoP/PhoQ Constituyen una Red Genética Robusta y Flexible

La habilidad de la arquitectura para reproducir diversos patrones es una propiedad sistémica, relativamente independiente de los parámetros escogidos. Como se observa en la Fig.3.12 para los constraints individuales, sólo una pequeña proporción de los parámetros del modelo está limitada a un subrango (25 %) de valores dentro del intervalo propuesto como biológicamente significativo. Si se reduce este rango al 5 %, no se observan parámetros que tengan una variabilidad menor a ese valor. Adicionalmente, una gran cantidad de parámetros (Fig. 3.13) puede tomar valores dentro de todo el rango propuesto (una variabilidad que se encuentra mucho más allá de lo esperable biológicamente). En los resultados de Robustez con respecto a los parámetros en las redes flexibles expuestos en la Figura 3.14 puede observarse que la robustez con respecto a la variación de parámetros disminuye, lo cual es coherente dado que se pide que la red realice funciones contrapuestas según el estímulo. Al reducir el rango al 5 %, se observa que al menos la mitad de los parámetros puede tener esa variabilidad, lo cual en la mayoría de los casos implica variaciones de al menos un orden de magnitud.

En conclusión, la habilidad de la red para reproducir patrones no se ve afectada por los parámetros individuales seleccionados, y estos pueden tomar valores dentro de un entorno de dimensiones significativas, con variaciones como las vistas en la Fig. 3.9 (nótese la gráfica de parámetros como H_PMRB). Desde el punto de vista computacional, esta robustez agrega una gran cantidad de complejidad al proceso de búsqueda de conjuntos de parámetros óptimos para la red. Sin embargo, desde el punto de vista biológico, este comportamiento puede evidenciar un fenotipo que confiere una ventaja adaptativa a lo largo de la evolución. Efectivamente, esta robustez podría permitir por un lado el mantenimiento del fenotipo global deseado a pesar de pequeños cambios (mutaciones) en el genotipo que determina la bioquímica de las especies involucradas. Por otro lado, la acumulación de pequeñas mutaciones puntuales permitiría, al transcurrir la evolución, que la red “navegara” aleatoriamente el espacio de valores posible sin afectar el fenotipo inicial, hasta alcanzar regiones en las cuales podrían convivir distintos patrones de funcionamiento (la flexibilidad observada en la red). Esto indica que el poder de la arquitectura para reproducir los patrones pedidos reside más en su topología (es decir, las especies y relaciones propuestas)

que en los valores de los parámetros escogidos.

4.4. La Autoregulación Positiva Amplifica la Señal Transducida

La autoregulación positiva de PhoP y PmrA a los operones PhoP/PhoQ y PmrA/PmrB posibilita la transducción de la señal de bajo Mg^{2+} y alto Fe^{3+} . Como se observa en los cuadros de resultados para los distintos modelos, no se observan modelos que pasen todos los tests requeridos en los cuales no se cuenta con autoregulación positiva de PhoP y PmrA. Esta autoregulación constituye un submódulo de feedback positivo como los descritos en [MSOI⁺02],[LR02], y constituye un submódulo dentro de los dos submódulos del sistema en estudio. Su inclusión es coherente con la hipótesis de que puede construirse una red funcional en base a pequeñas componentes básicas. En este caso, la funcionalidad agregada sería la de amplificar la señal a transducir.

4.5. La Transferencia de Información Es Soportada por la Fosforilación

Dentro del modelo propuesto, la transducción de señales se basa en las formas fosforiladas de PhoP y PmrA. En la sección anterior, al igual que en la información experimental publicada, no se hace mención explícita a si las especies consideradas (PmrA y PhoP) corresponden a las formas fosforiladas o no. En nuestro esquema de análisis (cf. modelos Reducido y Enriquecido) se exploran ambas variantes, concluyéndose que las especies responsables del feedback positivo son las fosforiladas. Esto resulta coherente con la función de estas subcomponentes, dado que si las especies no fosforiladas fueran las responsables del feedback (y considerando que la evidencia experimental propone un cierto nivel basal de actividad kinasa para PhoQ y PmrB) el aumento de la concentración de PhoP y PmrA podría amplificar una señal inexistente, contrariamente a la función de la red.

4.6. PmrA/PmrB y PhoP/PhoQ Presentan una Funcionalidad Coordinada

Las conexiones de ida y vuelta entre ambos sistemas de dos componentes resultan necesarias para la interacción de los mismos. En las distintas arquitecturas propuestas, se testearon opciones referentes al mecanismo de interacción entre los subsistemas PmrA/PmrB y PhoP/PhoQ. La conexión desde PmrA/PmrB a PhoP/PhoQ, que permite la inhibición de pmrD en presencia de alto Fe^{3+} , se modela a nivel transcripcional proponiendo a

PmrA-P y la asociación entre PmrD y PmrA-P como inhibidores de la transcripción, y la conexión desde PhoP/PhoQ hasta PmrA/PmrB, que permite la expresión de pbgP en `bajo Mg²⁺ independientemente de los niveles de Fe³⁺, se modela a nivel post-transcripcional como una asociación entre PmrD y PmrA-P. En todos los casos en los cuales no se incluyeron estas conexiones, las arquitecturas obtenidas perdieron su capacidad de reproducir los patrones biológicos. Es decir, no sólo estas condiciones permiten explicar la interacción entre los subsistemas, sino que su ausencia impide el funcionamiento de las redes según los criterios utilizados.

La propuesta de una interacción de ida a nivel post-transcripcional, involucrando a PmrD y a la especie fosforilada PmrA-P, entre los dos subsistemas de la red es coherente con la función biológica de la red genética, si se considera que son las especies fosforiladas las encargadas de activar a los genes controlados por PmrA. Efectivamente, una mayor transcripción de pmrA no garantizaría la existencia de una mayor concentración de la especie fosforilada de PmrA en ausencia del estímulo externo de alto Fe³⁺, y por lo tanto no aumentaría la expresión de los genes regulados por PmrA. Puede razonarse también que una actividad a nivel post-transcripcional aumenta la velocidad de respuesta del sistema al estímulo con respecto a la velocidad asociada a un evento transcripcional.

4.7. La Agregación de los Módulos PmrA/PmrB y PhoP/PhoQ Genera una Red Genética

Las interacciones entre los módulos PmrA/PmrB y PhoP/PhoQ determinan una red genética constituida por dos subsistemas. Los modelos implementados para la interacción entre los módulos PmrA/PmrB y PhoP/PhoQ permitieron especificar muchos puntos no considerados explícitamente hasta el momento. Esto, sumado a un análisis funcional que aborda exhaustivamente el comportamiento de la red, permite aportar evidencia desde una nueva óptica para sustentar (y eventualmente redirigir) la investigación “in vivo”.

Resulta claro que modelos tales como el mostrado en la Fig. 3.8 constituyen una hipótesis para el funcionamiento de la red, y debe considerarse que cada una de las conexiones entre los nodos (ejes numerados en la red) constituye una sub-hipótesis acerca de la interacción de los mismos (es decir, entre las especies). Asimismo, cada una de estas conexiones se implementa en el modelo mediante una cierta ecuación diferencial (Cuadro 3.21 y ecuaciones diferenciales), escogiendo entre distintas posibilidades como se describe en [MMOD02] y [Kae03]. El modelo computacional permite comprobar el funcionamiento de la red considerando la arquitectura como una agregación de sub-hipótesis “completa”, en el sentido de que se consideran todos los elementos necesarios para su funcionamiento. Esto contrasta con la experi-

mentación "in vivo" en la cual, tanto por costos como por desconocimiento de los mecanismos puntuales involucrados, se evalúa al sistema como una agregación de ciertos componentes escogidos sin formular hipótesis formales sobre el funcionamiento de las partes no evaluadas.

Las hipótesis de arquitectura para la red no tendrían valor si no estuvieran acompañadas por hipótesis de funcionamiento (como ya se ha expuesto, valores de concentración iniciales y finales para ciertas especies). Si bien el funcionamiento observado experimentalmente para la red es conocido, al formular todos los patrones esperables para la propuesta interacción de los subsistemas se consigue una visión a nivel de red del problema, permitiendo poner a prueba todas las componentes del sistema en un mismo experimento. Por lo tanto, la suma de los modelos presentados, la evidencia experimental analizada y los resultados obtenidos (determinando el realismo, flexibilidad y robustez del sistema) muestra que la funcionalidad de la red puede explicarse en base a la interacción de dos módulos interconectados tanto a nivel transcripcional como post-transcripcional.

4.8. La Activación Coordinada de pbgP Depende de la Sinergia de Mecanismos Transcripcionales y Post-transcripcionales

El modelo Enriquecido implementa una solución para interrogantes abiertos por el modelo Reducido. En este último se implementan los requerimientos considerados mínimos para el funcionamiento de la red. Sin embargo, en base a los resultados obtenidos (Cuadro 3.23) y a un análisis minucioso de las ecuaciones involucradas resulta claro que es necesario introducir una nueva hipótesis alrededor de la conexión de ida vía PmrD entre los dos subsistemas.

Siguiendo con la nomenclatura de los ejes de la Figura 3.3, puede observarse que la acción de PmrD consiste en asociarse con PmrA-P (17) para conformar un dímero que puede activar la transcripción de pbgP (20) e inhibir la de pmrD (24). PmrB en su configuración normal no afecta a PmrA, pero en presencia de Fe³⁺ cambia a un estado "activado" (3), en el cual puede tanto fosforilar como defosforilar a PmrA y PmrA-P, respectivamente (11 y 12). Cómo los parámetros de las ecuaciones derivadas de los ejes 11 y 12 son distintos, existe la posibilidad de que adopten valores que hagan que la reacción 11 sea más eficiente que la 12, con el consiguiente aumento de la fosforilación de PmrA en presencia de Fe³⁺. Ahora bien, este mecanismo plantea un problema si se requiere que la red active la transcripción de pbgP con Mg=1 y Fe=0, dado que esta activación (19,20) se produce utilizando PmrA-P, el cual no estará disponible al no existir PmrB-ACT en ausencia de Fe³⁺.

En el modelo Enriquecido (Fig. 3.8) se propone una hipótesis alternativa para la interacción entre PmrB y PmrA (y análogamente, PhoQ y PhoP)

que permite explicar cómo el sistema real soluciona el problema expuesto. Se propone que PmrB en su estado normal posee tanto actividad kinasa (12.2) como fosfatasa (12.1) y que en presencia de Fe^{3+} el estado activado de PmrB, PmrB-ACT, presenta sólo actividad kinasa (11). Análogamente a lo propuesto en el modelo Reducido, las ecuaciones 12.1 y 12.2 dependen de distintos dominios activos dentro de la proteína PmrB y por lo tanto sus cinéticas no son necesariamente iguales, por lo que se utilizan ecuaciones con distintos parámetros para describirlas. Al ser los parámetros distintos, pueden tomar distintos valores que permitan diferencias entre las reacciones, posibilitando por ejemplo que exista una cierta cantidad basal de PmrA-P, indispensable para la interacción con PmrD. Los resultados obtenidos con respecto al realismo, flexibilidad y robustez del modelo Enriquecido, así como al comparación con resultados experimentales (Cuadro 3.29) muestran que este es un modelo factible, y abren las puertas para la realización de nuevas pruebas de laboratorio encaminadas a verificar sus distintas sub-hipótesis.

La hipótesis descrita permite describir un mecanismo de activación para pbgP basado en una cierta cantidad basal de PmrA-P, en presencia de bajo Mg^{2+} y bajo Fe^{3+} . Este hecho hace necesaria la formulación de una nueva hipótesis referente a la fuerza de esta activación. Efectivamente, es necesario explicar el hecho de que, existiendo una cantidad basal de PmrA-P (único activador de la transcripción de pbgP según la evidencia de binding sites) pbgP no se encuentra basalmente activado.

Este camino no se estudia explícitamente en este trabajo, pero los resultados obtenidos permiten formular hipótesis a analizar en futuros trabajos. Dado la activación de pbgP depende tanto de PmrA-P como de PmrA-P_PmrD, y se postula que debe existir una cierta cantidad basal de PmrA-P, una posibilidad es que pbgP se active sólo al superarse una cierta concentración de PmrA-P (superior a la basal), independientemente de si las especies PmrA-P y PmrA-P_PmrD tiene distintas afinidades por el promotor de pbgP o no.

Siguiendo con esta línea de razonamiento, otra posibilidad es que el dímero PmrA-P_PmrD tenga una afinidad mayor por el promotor de pbgP que la que posee PmrA-P, de forma tal que la activación de pbgP se lograría en caso de aumentar considerablemente la concentración de PmrA-P (en condiciones de alto Fe^{3+}), o en caso de que una parte de este PmrA-P basal se asociara con PmrD para formar un dímero de mayor afinidad por el promotor (condiciones de bajo Fe^{3+} y bajo Mg^{2+}).

4.9. GENIE Predice la Existencia de Grupos de Genes Regulados por PHOP

En base a la validación experimental realizada en la sección 3.5, puede concluirse que GENIE permite realizar predicciones sobre la actividad de

distintos genes regulados por la red genética estudiada. Efectivamente, una vez obtenida (aprendida) una solución válida para el sistema estudiado, resulta posible modificar los valores de los parámetros del sistema dentro de un cierto entorno numérico de la misma, es decir modificar en un cierto rango predefinido sólo los parámetros asociados a la especie cuyo comportamiento se intenta modificar, para predecir, por ejemplo, los patrones de activación de nuevos genes. En el caso estudiado de los genes regulados por PhoP, los parámetros modificados inciden directamente sobre el tiempo de activación, activación máxima y tiempo de inicio de la disminución de la expresión. Estas modificaciones predicen la cinética de genes cuya expresión varía en función de estas características, clasificándolos en forma automática. La existencia de estos grupos de genes predichos se comprueba al analizar el comportamiento de los genes regulados por PhoP en los experimentos in vivo.

4.10. GENIE Permite Generar Hipótesis Adaptativamente

El enfoque utilizado permite ajustar adaptativamente las arquitecturas propuestas en función de la evidencia acumulada. Efectivamente, las distintas iteraciones efectuadas durante el análisis del problema biológico derivaron en nuevas hipótesis que fueron agregadas a la arquitectura propuesta. Algunas de estas hipótesis fueron incluidas exitosamente en el modelo, y otras fueron descartadas, aportando ambos casos importante información para el análisis. El enfoque utilizado para la adición de hipótesis al modelo se basó en considerar a la red como la interacción de distintos módulos funcionales. Desde esta óptica, la red puede visualizarse como la interacción de dos subsistemas (los dos sistemas de dos componentes PmrA/PmrB y PhoP/PhoQ) relativamente análogos. A su vez, cada subsistema fue considerado como la interacción de distintos motivos, como por ejemplo la autoregulación positiva de los genes *phoP* y *pmrA* mediante las especies fosforiladas PhoP-P y PmrA-P introducida a partir del modelo Alternativa 2.1. Esta visión permitió realizar experimentos sobre la red explorando distintas combinaciones entre motivos, y asimismo introducir motivos novedosos como la interacción a nivel post-transcripcional vía PmrD. Este último motivo surgió como hipótesis de trabajo en base a la evidencia acumulada hasta esa etapa del análisis y a la interacción con el grupo de investigación del Groisman Lab. (Washington University in St. Louis, USA).

La toma de decisiones con respecto a agregar o no elementos a la arquitectura fue posible gracias a la definición de distintas condiciones a cumplir por la red obtenida, como realismo, flexibilidad y robustez. Todas estas condiciones son cumplidas por el último modelo obtenido (Enriquecido).

El enfoque propuesto para la modelización de redes genéticas provee un

marco conceptual de trabajo, que incorpora elementos bioquímicos como las distintas ecuaciones posibles para el modelado de las reacciones, biológicos como las distintas funciones de la red a reproducir y computacionales, como el uso de simulaciones sobre los modelos, la exploración del espacio de soluciones y la definición de características sistémicas a cumplir por la red, entre otras. Queda entonces definida una base tanto para el análisis sistemático de nuevas redes genéticas, como para la optimización de las distintas componentes computacionales del modelo, permitiendo acrecentar el poder predictivo del mismo y aumentar el grado de automatización de los procesos.

4.11. Trabajos Futuros

La metodología de trabajo GENIE presentada en este trabajo se encuentra constituida por una agregación de distintos módulos, sobre los cuales pueden realizarse optimizaciones y extensiones.

El mecanismo de búsqueda de soluciones utilizado por GENIE constituye una primera aproximación al problema, y podría ser sustituido por métodos de búsqueda más refinados, como por ejemplo algoritmos genéticos, que tomen ventaja de las características especiales del problema modelado. Una vez introducidos estos nuevos métodos de búsqueda en el modelo, resultaría de interés encontrar implementaciones de gran eficiencia, por ejemplo mediante la paralelización de los algoritmos, que permitieran disminuir el tiempo de computación necesario para procesar el algoritmo, derivado de la gran dimensionalidad del problema.

La integración entre la inferencia de interacciones genéticas basada en información de binding sites y la generación de modelos dinámicos realizada por GENIE se encuentra en un estado embrionario, siendo posible realizar desarrollos en este sentido que proporcionen una implementación automática del modelo teórico. En este sentido, podrían investigarse nuevos métodos para la regeneración de hipótesis que optimicen la forma en la que GENIE explora las arquitecturas posibles para un conjunto de reglas sobre el funcionamiento del sistema biológico.

Finalmente, GENIE aporta elementos para la automatización del proceso de *drug discovery* sobre la red genética estudiada. En este sentido, podría ser objeto de futuras investigaciones el desarrollo de nuevos módulos GENIE para la generación de predicciones orientada a esta fin.

Bibliografía

- [BG03a] E. Batchelor and M. Goulian. Robustness and the cycle of phosphorylation and dephosphorylation in a two-component regulatory system. *PNAS*, 100(2):691–696, Jan 2003.
- [BG03b] E. Batchelor and M. Goulian. Supporting text: Robustness and the cycle of phosphorylation and dephosphorylation in a two-component regulatory system. *PNAS*, 100(2):691–696, Jan 2003.
- [Bre00] S. Brenner. The end of the beginning. *Science*, (287):2173–2174, Mar 2000.
- [CEHL02] Guet C.C, M.B. Elowitz, W. Hsing, and Stanislas Leibler. Combinatorial synthesis of genetic networks. *Science*, 296(5572):1466–70, May 2002.
- [DZG03] S. Dongwoo, I. Zwir, and E. Groisman. Positive autoregulation of the two-component system phop/phoq triggers sequential activation of the phop regulon. 2003.
- [Esp01] Real Academia Española. *Diccionario de la lengua española*. 22 edition, 2001.
- [FL98] S. Fuhrman and S. Liang. Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Pacific Symposium on Biocomputing*, pages 18–29, 1998.
- [GHHS95] E. Groisman, J. Heffron, J.A. Hoch, and T.J. Silhavy. Two-component signal transduction. *Am. Soc. Microbiol. Press.*, 1995.
- [HMC02] J. Hasty, D. McMillen, and J. Collins. Engineered gene circuits. *Nature*, 420(6912):224–230, Noviembre 2002.
- [Kae03] M. Kaern. *Regulatory dynamics in engineered gene networks*. Boston University, Centre for BioDynamics, 2003.

- [KG04] A. Kato and E. Groisman. A protein shield that protects a response regulator from dephosphorylation by its cognate sensor. 2004.
- [KLG03] A. Kato, T. Latifi, and E. Groisman. Closing the loop: The pmra/pmrb two-component system negatively controls expression of its posttranscriptional activator pmrd. *PNAS*, 100:4706–4711, 2003.
- [LR02] T. Lee and N. Rinaldi. Transcriptional regulatory networks in *saccharomyces cerevisiae*. *Science*, 298(5594):799–804, Oct 2002.
- [LW01a] C. Li and W. Wong. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *PNAS*, 98(1):31–36, Jan 2001.
- [LW01b] C. Li and W. Wong. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biology*, 2(8):RESEARCH0032, Aug 2001.
- [MA99] H. McAdams and A. Arkin. It's a noisy business. *Trends Genet*, 15(2):65–69, Feb 1999.
- [MMCW00] E. Mjolsness, T. Mann, R. Castaño, and B. Wold. From coexpression to coregulation: an approach to inferring transcriptional regulation among gene classes from large-scale expression data. *Advances in neural information processing systems*, 12:928–934, 2000.
- [MMOD02] E. Meir, E. Munro, G. Odell, and G. Von Dassow. Ingeneue: a versatile tool for reconstructing genetic networks, with examples from the segment polarity network. *Journal of experimental zoology*, 294(3):216–51, Oct 2002.
- [MSOI⁺02] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskil, and U. Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, Octubre 2002.
- [MvDMO02a] E. Meir, G. von Dassow, E. Munro, and G.M. Odell. Robustness, flexibility, and the role of lateral inhibition in the neurogenic network. *Current Biology*, 12(10):778–786, May 2002.
- [MvDMO02b] E. Meir, G. von Dassow, E. Munro, and G.M. Odell. Robustness, flexibility, and the role of lateral inhibition in the neurogenic network. - supplementary material. 2002.

- [vDMMO00a] G. von Dassow, E. Meir, E.M. Munro, and G.M. Odell. Formulation of a model of the segment polarity network as a system of first-order ordinary differential equations using in-geneue. *Nature*, 406(6792):188–192, Jul 2000.
- [vDMMO00b] G. von Dassow, E. Meir, E.M. Munro, and G.M. Odell. The segment polarity network is a robust developmental module. *Nature*, 406(6792):188–192, Jul 2000.
- [vSWEB02] E.P. van Someren, L.F.A. Wessels, and M.J.T. Reinders. E. Backer. Genetic network modeling. *Pharmacogenomics*, 3(4):507–525, Jul 2002.
- [WHA] M. Wahde, J. Hertz, and M.L. Andersson. Reverse engineering of sparsely connected genetic regulatory networks.
- [ZTG03] I. Zwir, P. Traverso, and E. Groisman. Semantic-oriented analysis of regulation: the phop regulon as a model network. *International Conference on Systems Biology*, 2003.

