

Universidad de Cantabria
Facultad de Medicina
Departamento de Biología Molecular



**“Parametrización de la regulación
transcripcional en bacterias:
TetR como paradigma”**

Memoria presentada para optar al grado de Doctor por
Raúl Ruiz González

El presente trabajo ha sido realizado en el Departamento de Biología Molecular de la Universidad de Cantabria y el Instituto de Biomedicina y Biotecnología de Cantabria, bajo la dirección del catedrático Fernando de la Cruz Calahorra y el doctor Raúl Fernández López, gracias a una beca predoctoral FPI (Formación de Personal Investigador) concedida por el Ministerio Ciencia e Innovación de referencia BES-2009-013371.

D. Fernando de la Cruz Calahorra, Catedrático de Genética, y D. Raúl Fernández López, investigador postdoctoral, pertenecientes al departamento de Biología Molecular de la Universidad de Cantabria,

CERTIFICAN:

Que D. Raúl Ruiz González ha realizado bajo su dirección el trabajo que lleva por título “Parametrización de la regulación transcripcional en bacterias: TetR como paradigma”.

Consideramos que dicho trabajo se encuentra terminado y reúne los requisitos necesarios para su presentación como memoria de Doctorado, al objeto de poder optar al grado de Doctor por la Universidad de Cantabria

En Santander, a 25 de enero de 2016.

Fdo. Fernando de la Cruz Calahorra

Fdo. Raúl Fernández López

A mi familia

Agradecimientos

En primer lugar me gustaría agradecer a mi director y codirector la realización de esta tesis. Fernando, gracias por darme la oportunidad de trabajar en tu laboratorio y confiar en mí durante estos años. Raúl, gracias por todo lo que me has enseñado y porque sin ti esta tesis no existiría.

En segundo lugar me gustaría dedicar unas palabras a dos personas que me han influido y acompañado durante estos años. Por un lado, Mapi, gracias por cuidar de mí y ayudarme siempre que lo he necesitado. Y por otro, Alejandro, gracias por ser mi amigo y haber creído en mi más que yo mismo.

En tercer lugar me gustaría dar las gracias a Gabi, Iñaki y Elena y sus grupos. Siempre me ha parecido beneficioso colaborar tan estrechamente y tener diferentes puntos de vista a la hora de discutir nuestra investigación.

Por último, pero no menos importante, me gustaría agradecer a todos los compañeros de laboratorio que han compartido conmigo este periodo. He aprendido mucho de ellos y espero haberles sido de ayuda también. Irene y Ana, que me ayudaron a aterrizar en el laboratorio. A Sheila, Carlos y Raquel, por ser tan buenos compañeros. A Atha y María, mis postdocs de cabecera. Gracias a la vieja guardia: Jorge, Esther, David, Val, Inma, Juan, Yera, Andrés... Y también a la siguiente: María, Maris, Coral... Hemos pasado unos momentos geniales. Espero que los recién llegados tengan tanta suerte como he tenido yo. Durante estos años muchas más personas han compartido conmigo su tiempo y han ayudado a la finalización de esta tesis:

Gracias a todos

Contenido

Abreviaturas	15
1. Introducción	19
1.1. Transcripción.....	19
1.2. Reguladores transcripcionales.....	20
1.3. TetR y su familia de reguladores transcripcionales	23
1.3.1. Estructura y mecanismo general de la familia TetR	23
1.3.2. Mecanismo de regulación de TetR.....	25
1.4. Regulación analógica frente a regulación digital	26
1.5. Funciones de transferencia.....	27
1.5.1. Métodos experimentales para medir la función de transferencia	29
1.5.2. Estocasticidad en la expresión genética y sus consecuencias en la función de transferencia	30
1.5.3. Función de transferencia y toma de decisiones.....	32
1.6. Entropía de una distribución	33
1.7. Información mutua	34
1.7.1. Uso de la información mutua en la regulación transcripcional	35
Objetivos:	41
2. Materiales y Métodos	45
2.1. Cepas.....	45
2.2. Cultivo	45
2.2.1. Medios de cultivo	45
2.2.2. Antibióticos y correguladores	46
2.3. Oligonucleótidos	46
2.4. Plásmidos	48
2.5. Métodos de clonaje	49
2.5.1. Extracción y purificación de DNA	49
2.5.2. Cuantificación del DNA.....	50
2.5.3. Reacción en cadena de la polimerasa (PCR)	50
2.5.4. Electroforesis de DNA en gel de agarosa	51
2.5.5. Digestión de DNA mediante enzimas de restricción	51
2.5.6. Fosforilación y desfosforilación.....	52
2.5.7. Reacciones de ligación	52
2.5.8. Transformación por electroporación	52

2.5.9.	Método de ensamblaje isotérmico (<i>Isothermal assembly</i>).....	53
2.5.10.	Recombinación cromosómica con productos de PCR: método Datsenko-Wanner	55
2.5.11.	Mutación sitio específica.....	57
2.5.12.	Secuenciación de DNA.....	58
2.6.	Construcción de dispositivos para el análisis de la regulación transcripcional	58
2.7.	Cuantificación de la actividad transcripcional mediante fluorímetro	62
2.8.	Cuantificación de la actividad transcripcional mediante microscopía de fluorescencia	63
2.8.1.	Adquisición de imágenes.....	63
2.8.2.	Tratamiento de imágenes	64
2.9.	Simulaciones y análisis computacional	68
2.9.1.	Ecuaciones maestras	68
2.9.2.	Procesos de Markov y algoritmo de Gillespie	71
2.9.3.	Medida de la información mutua (I):	80
3.	Resultados	87
3.1.	Cuantificación de la regulación transcripcional en cultivo mediante fluorímetro	87
3.1.1.	Actividad transcripcional del promotor	87
3.1.2.	Caracterización de los sistemas de expresión.....	88
3.1.3.	Actividad reguladora de TetR.....	90
3.1.4.	Fusión TetR:mKate2	93
3.1.5.	Dispositivo genético de medida de un solo vector e inserción cromosómica	95
3.1.6.	Control de las proteínas fluorescentes: TetR:GFP y p _{tetA} :mKate2	99
3.2.	Cuantificación de la regulación transcripcional mediante microscopía de epifluorescencia	100
3.2.1.	Sistema de dos vectores: pRRG13 y pRRG62 en BW27783	101
3.2.2.	Sistema de un vector: pRRG63 en BW27783	105
3.2.3.	Inserción cromosómica del sistema: cepa RRS247	109
3.3.	Capacidad informativa del sistema regulador: cálculo de la información mutua	114
3.4.	Análisis de los factores que modifican la información mutua en un circuito transcripcional.....	118
3.4.1.	La DPI en un circuito transcripcional.....	120
3.4.2.	Influencia de la vida media de los componentes en la información mutua.	123
3.4.3.	Efecto de las fluctuaciones sobre la información mutua.	127
3.4.4.	Efecto de K y n sobre la información mutua	131

4. Discusión	137
4.1. El problema de la parametrización de las redes de regulación transcripcionales	137
4.2. La fluorimetría de cultivos presenta problemas de sensibilidad a la hora de determinar la función de transferencia.	138
4.3. La microscopía de fluorescencia permite resolver el problema de la sensibilidad experimental	140
4.4. El comportamiento de la media y la mediana frente a la dispersión poblacional	140
4.5. Diferencias en el rango de la función de transferencia debidas a la localización genética del sistema.....	141
4.6. Efecto de la localización genética del sistema sobre las fluctuaciones	142
4.7. Ajustes de la función de transferencia a la mediana poblacional: ¿una o dos fases?.....	144
4.8. Divergencia entre los datos bioquímicos y los resultados in vivo	146
4.9. Control analógico frente a control digital.....	147
4.10. Información mutua en la regulación por TetR.....	149
4.11. TetR actúa como un controlador digital ruidoso	150
4.12. La desigualdad en el procesamiento de la información.....	151
4.13. Efecto de la velocidad de degradación en la información mutua	152
4.14. La información transmitida por TetR frente a otros reguladores transcripcionales .	152
4.15. Optimización de la transmisión de información.....	154
Conclusiones	159
Bibliografía	163

Abreviaturas

Amp	Ampicilina
aTc	Anhidrotetraciclina
ATP	Adenosín trifosfato
cAMP	Adenosín monofosfato cíclico
Cm	Cloranfenicol
CFP	Proteína fluorescente cian
CME	Ecuación maestra
CRP	Proteína de unión a cAMP
C-terminal	Carboxi-terminal
dATP	Desoxiadenosina trifosfato
DBD	Dominio de unión al DNA
dCTP	Desoxicitidina trifosfato
dGTP	Desoxiguanosina trifosfato
DJ	Desigualdad de Jensen
DTT	Ditiotreitol
dTTP	Desoxitimidina trifosfato
DNA	Ácido desoxirribonucleico
DPI	Desigualdad en el procesado de la información
EBD	Dominio de unión al efector
GFP	Proteína fluorescente verde
HTH	Motivo estructural <i>helix-turn-helix</i>
IPTG	Isopropil- β -D-tiogalactopiranosido
IR	Repetición invertida
kb	Kilobase
Km	Kanamicina
LB	Medio de cultivo Luria-Bertani Broth o Lysogeny Broth
MCS	Sitio múltiple de clonaje
mRNA	Ácido ribonucleico mensajero
NAD	Nicotinamida adenina dinucleótido
N-terminal	Amino-terminal
OD	Absorbancia o densidad óptica
ODE	Ecuación diferencial ordinaria
bp	Pares de bases
PCR	Reacción en cadena de la polimerasa
PEG	Polietilenglicol
p/v	Relación peso/volumen
RHH	Motivo estructural <i>ribbon-helix-helix</i>
RNA	Ácido ribonucleico
rpm	Revoluciones por minuto
RTT	Reguladores de tipo TetR
SOB	Medio de cultivo Super Optimal Broth
TF	Factor de transcripción
Tc	Tetraciclina
u.a.f.	Unidades arbitrarias de fluorescencia
v/v	Relación volumen/volumen
wHTH	Motivo estructural <i>winged helix-turn-helix</i>
YFP	Proteína fluorescente amarilla
3D	Tridimensional

Introducción

1. Introducción

1.1. Transcripción

Los organismos codifican la información necesaria para ejecutar sus funciones vitales en el DNA. Aunque el DNA de una célula contiene la información necesaria para la síntesis de todas las proteínas enzimáticas y estructurales que necesita, no todas estas proteínas se sintetizan simultáneamente. La célula regula la expresión de su genoma de acuerdo a señales internas y externas, por lo que el fenotipo de un organismo está críticamente condicionado por la regulación de la expresión de su genoma.

En bacterias, la regulación de la expresión genética se realiza fundamentalmente a nivel de la transcripción. La transcripción es el proceso por el cual la cadena de DNA se transcribe en un intermediario de RNA, que después será traducido a proteína. La transcripción abarca cuatro etapas (Lewin, 1994):

- i) Reconocimiento del DNA molde (ver Figura 1-1): unión de la RNA polimerasa al DNA en la región promotora y formación de la burbuja de transcripción mediante la separación de las cadenas alrededor del sitio de unión de la RNA polimerasa.
- ii) Iniciación (Figura 1-1): síntesis de los primeros enlaces entre ribonucleótidos. El holoenzima ocupa las posiciones -55 a +20 del gen y permanece en el promotor hasta que sintetiza aproximadamente las primeras 9 bases. Si el transcrito inicial no consigue esa longitud es liberado y el enzima comienza nuevamente la síntesis.
- iii) Elongación: el enzima se mueve a lo largo del DNA y extiende la cadena de RNA. A medida que se mueve el enzima, el DNA se desenrolla y un nuevo fragmento de la cadena molde queda expuesto en forma de cadena sencilla. En la parte desenrollada se encuentra el híbrido RNA-DNA y detrás de éste se recompone la doble hélice de DNA, mientras que el RNA emerge como cadena sencilla. Una vez que comienza la elongación, el holoenzima pierde los contactos con la región -55 a -35 y abarca una región de unos 60 bp. Cuando el transcrito alcanza unas 25-30 bases, el núcleo del enzima sufre una transición para continuar la elongación y solo abarca unos 30 bp.
- iv) Terminación: reconocimiento del sitio a partir del cual no se pueden adicionar más bases a la cadena de RNA. La burbuja colapsa, el híbrido RNA-DNA se deshace y tanto el enzima como el RNA son liberados del DNA. Estas reacciones necesitan de una secuencia

del DNA llamada terminador transcripcional.

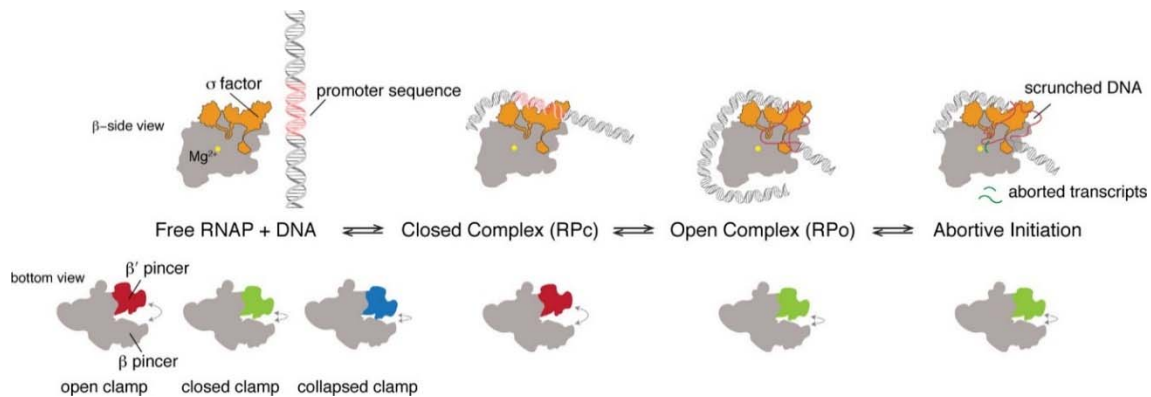


Figura 1-1. Resumen de los pasos en la iniciación de la transcripción en procariotas. La abrazadera de la RNA polimerasa libre puede existir en tres conformaciones diferentes: abierta (coloreada en rojo), cerrada (verde) y colapsada (azul). El holoenzima interactúa con el promotor a través del factor σ para formar el complejo promotor cerrado. En este estado solo se ha observado la abrazadera en su conformación abierta. En el complejo promotor abierto, la RNA polimerasa enrolla alrededor de sí misma el DNA de la región ubicada aguas arriba del promotor, cierra la abrazadera y desenrolla el promotor. A continuación, la RNA polimerasa sintetiza y libera transcritos cortos en un proceso conocido como iniciación abortiva antes de pasar a la fase de elongación. Tomada de (Dangkulwanich *et al.*, 2014).

¿Por qué la transcripción evolucionó en un proceso celular esencial? ¿Por qué no se expresa directamente la información codificada en el genoma en proteínas? Hay varias razones que justifican la evolución de la transcripción como un paso intermedio para la síntesis de proteínas. Primero, la transcripción expande la variedad de productos génicos a través del mecanismo de corte y empalme (*splicing*). Segundo, copiando la información del DNA a muchas moléculas de RNA se incrementa la velocidad de síntesis de proteínas en la célula y se evita el cuello de botella que podría resultar de la expresión directa de un gen a partir del DNA. Tercero, el número de moléculas de RNA disponibles en cada momento para sintetizar proteínas puede ser regulado con precisión para dar lugar a una explosión de productos. La amplificación de la señal implícita en el proceso de transcripción incrementa el rango dinámico de la expresión, permitiendo a la célula controlar el producto traducido del RNA con más precisión y en una manera específica de gen. Esta amplificación también da lugar a la aleatoriedad en la expresión génica, haciendo posible producir varias respuestas a partir de células genéticamente idénticas (Kaern *et al.*, 2005).

1.2. Reguladores transcripcionales

Probablemente por razones de economía, las bacterias controlan la expresión de su genoma fundamentalmente a nivel del inicio de la transcripción. Para ello utilizan proteínas específicas capaces de unirse a la región promotora del gen, activando o reprimiendo su transcripción por la RNA polimerasa (Figura 1-2). Los factores de transcripción (TFs) (también llamados reguladores transcripcionales) pueden ser por tanto de dos tipos: activadores y represores de

la transcripción. Los represores transcripcionales se unen a los sitios operadores del promotor, bloqueando la asociación de la RNA polimerasa o previniendo su progresión. Con frecuencia, los represores transcripcionales son capaces de unirse a una molécula efectora que modifica su afinidad por los sitios operadores. Denominamos inductores a aquellos efectores que producen la liberación de la unión TF-sitio operador. Denominamos aporrepresores a aquellos TF que, al unirse a su efector (correpresor) específico, son capaces de unirse a sus sitios operadores. (Figura 1-2.B). Los activadores transcripcionales se unen a sus sitios operadores reclutando la RNA polimerasa al promotor o induciendo la formación de complejos abiertos promotor-RNA polimerasa transcripcionalmente activos cuando el efector está presente (Figura 1-2.C). La activación transcripcional es más compleja que la represión y requiere frecuentemente la curvatura del DNA y el establecimiento de contactos específicos con la subunidad α de la RNA polimerasa (Fernandez-Lopez *et al.*, 2015).

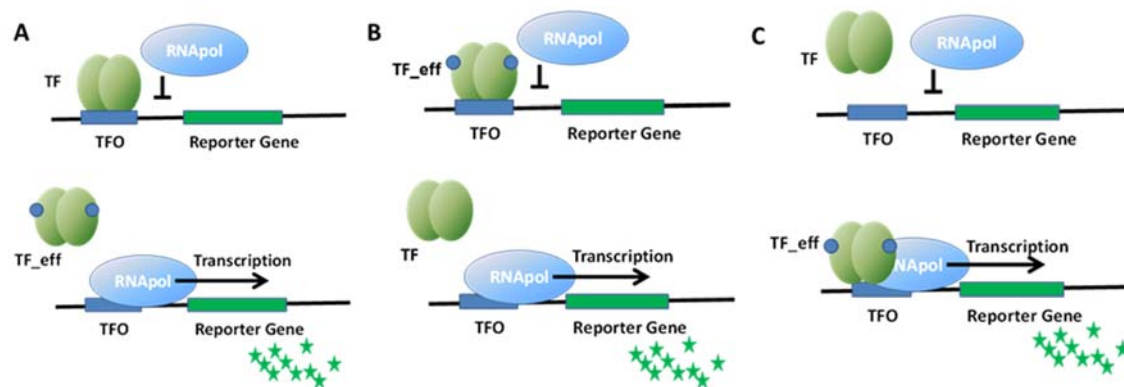


Figura 1-2. Representación esquemática de tres tipos de regulación transcripcional, basada en represores (A), aporrepresores (B), o activadores (C). La unión del factor transcripcional (TF, representado por óvalos verdes) a su sitio operador (TFO en la figura, representado por un rectángulo azul), localizado dentro de la región promotora, afecta la actividad de la RNA polimerasa (RNAPol en la figura y representada por un óvalo azul) y, por tanto, la señal asociada a la proteína regulada (estrellas verdes). La presencia de una molécula efectora (representada por círculos azules pequeños) modifica la interacción del factor transcripcional con su operador y cambia la cantidad de proteína reportera producida. Tomado de (Fernandez-Lopez *et al.*, 2015).

Los TFs contienen típicamente un dominio de unión a DNA (DBD, de *DNA binding domain*), responsable del reconocimiento y unión al operador, y otro dominio de unión al efector (EBD, de *effector binding domain*), que es responsable de la oligomerización del regulador y la transmisión de la señal efectora al dominio de unión al DNA. Los DBDs en bacterias muestran una gran conservación, existiendo tres tipos básicos de dominios: de tipo *helix-turn-helix* (HTH), de tipo *ribbon-helix-helix* (RHH) y los de hélice alada (wHTH de *winged helix-turn-helix*). Por el contrario, los EBDs bacterianos muestran una enorme diversidad, existiendo TFs capaces de unirse a metales, moléculas complejas o incluso responder a señales fisicoquímicas como el pH o la luz.

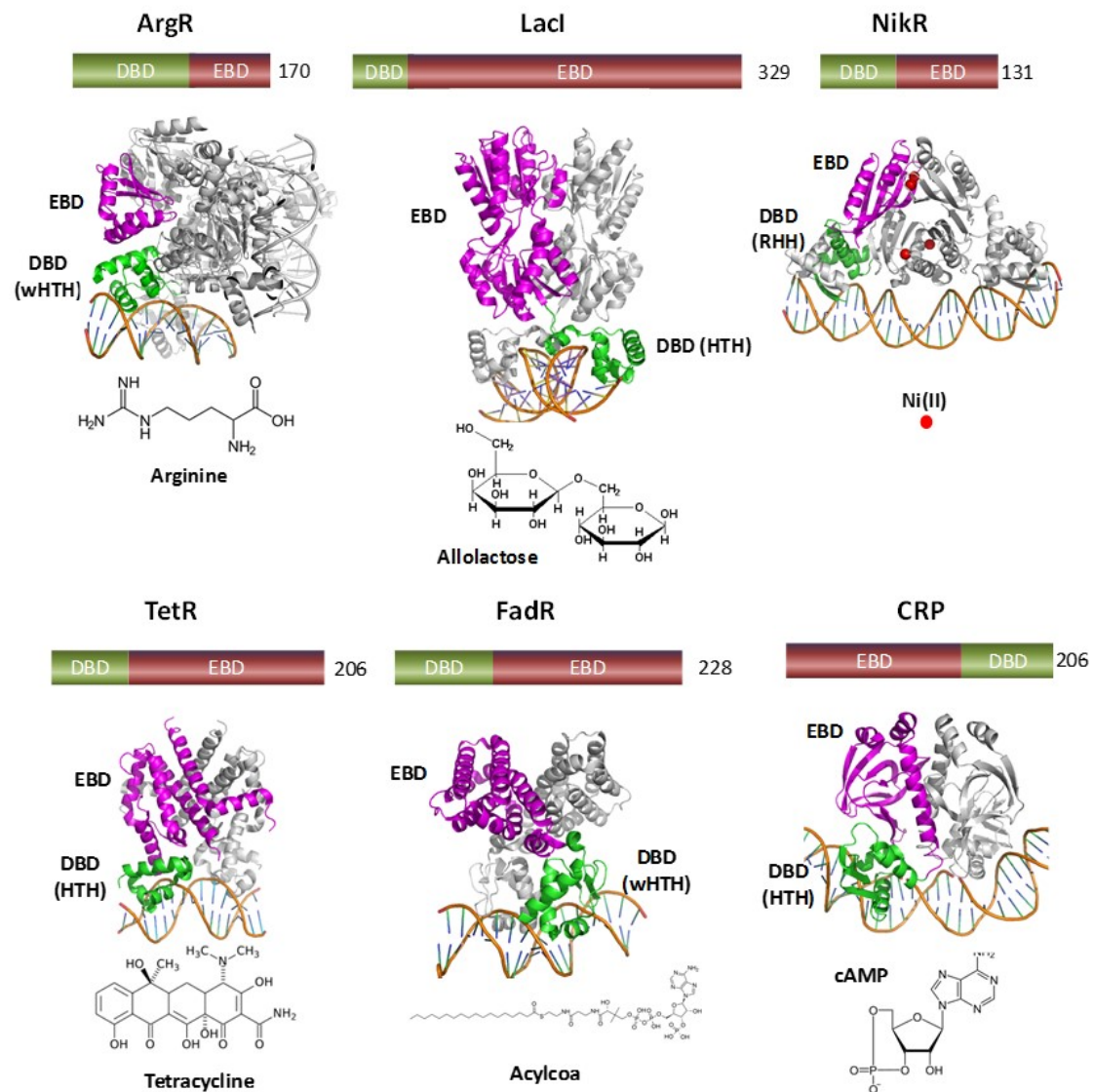


Figura 1-3. Variedad de factores transcripcionales de un componente presentes en la naturaleza. (A) ArgR (Número de acceso a RCSB Protein Data Bank 3LAJ, (Cherney *et al.*, 2010)) reprime la transcripción de los genes biosintéticos del operón arginina. ArgR es un aporrepresor transcripcional del tipo wHTH que se activa por el correpresor arginina, de manera que el operón no se transcribe cuando el efector está presente. (B) LacI (1EFA, (Bell and Lewis, 2000)) inhibe la expresión del operón *lac*. Cuando el azúcar alolactosa está presente, su unión al dominio C-terminal del dímero de LacI produce un cambio conformacional que inhibe la unión del DNA por parte del dominio N-terminal HTH. (C) NikR (2HZV, (Schreiter *et al.*, 2006)) es un aporrepresor dependiente de ligando que solo se une al operón cuando se encuentra unido al catión Ni^{+2} . (D) TetR (1QPI, (Orth *et al.*, 2000)) es un represor del tipo HTH que siente la presencia del antibiótico tetraciclina. (E) FadR (1HW2, (Xu *et al.*, 2001)) contiene un motivo N-terminal de tipo wHTH conectado a un dominio C-terminal EBD similar a los EBDs de la familia TetR. Sin embargo, los efectores de FadR son acil-CoAs en vez de antibióticos. (F) La proteína receptora de cAMP (1CGP, (Schultz *et al.*, 1991), conocida como CRP, es un activador transcripcional. La presencia de cAMP permite la unión del motivo C-terminal HTH de CRP al operador de varios operones catabólicos. Una vez unido al operador, CRP-cAMP interactúa con la RNA polimerasa, permitiendo la transcripción de los correspondientes operones. Los dominios DBD y EBD están coloreados en verde y magenta, respectivamente. Se indica el número de aminoácidos de cada factor de transcripción. Tomada de (Fernandez-Lopez *et al.*, 2015).

1.3. TetR y su familia de reguladores transcripcionales

TetR es el regulador de respuesta al antibiótico tetraciclina, un represor transcripcional que reprime la expresión del gen de resistencia *tetA*. TetR utiliza la tetraciclina como efector, de tal forma que la presencia de tetraciclina en el medio libera la unión de TetR al promotor de TetA (p_{tetA}), permitiendo la expresión de este gen de resistencia. TetR constituye uno de los modelos principales de regulación transcripcional por tres razones fundamentales:

- a) Debido a su papel en la resistencia a los antibióticos, TetR es uno de los reguladores mejor conocidos a nivel bioquímico. Las bases moleculares del reconocimiento tanto de la secuencia operadora *tetO* como del inductor (tetraciclina) se conocen con mucho detalle, existiendo distintos mutantes con afinidades diferentes y especificidades alteradas para el inductor (Helbl and Hillen, 1998; Stanton *et al.*, 2014).
- b) TetR puede utilizar como efectores análogos de la tetraciclina que no poseen actividad antibiótica, siendo el más utilizado la anhidrotetraciclina (aTc). Este hecho, unido al amplio conocimiento de su bioquímica y a la capacidad de TetR de funcionar tanto en células procariotas como eucariotas, han hecho de TetR una de las herramientas biotecnológicas más utilizadas. TetR y sus derivados se utilizan frecuentemente para la expresión controlada de genes en procesos biotecnológicos y en biología sintética (Berens and Hillen, 2004; Sotiropoulos and Kaznessis, 2007).
- c) TetR es el prototipo de una amplia familia de reguladores transcripcionales. Los reguladores de tipo TetR (RTT) constituyen uno de los sistemas regulatorios más comunes en bacterias. Los RTT están fundamentalmente asociados con la resistencia a antibióticos y la regulación de genes que codifican exportadores de pequeñas moléculas, aunque también regulan otras funciones celulares (Cuthbertson and Nodwell, 2013; Ramos *et al.*, 2005). Los RTT han sido encontrados en prácticamente todos los géneros de procariotas, normalmente con varios RTT por célula bacteriana.

1.3.1. Estructura y mecanismo general de la familia TetR

Estructuralmente, los RTT contienen un DBD N-terminal y un EBD C-terminal. El dominio DBD, formado por las hélices 1-3 está altamente conservado y contiene el motivo HTH. La hélice α lleva a cabo el reconocimiento, insertándose en el surco mayor del operador. Se han resuelto varias estructuras de RTT con sus respectivos operadores. El operador de RTT típico presenta una repetición invertida (IR) de 15 bp, con dos brazos de 6 bp separados por 1 bp. El motivo

1.3.2. Mecanismo de regulación de TetR

TetR es la proteína tipo que da nombre a la extensa familia de proteínas RTT. Con el nombre de TetR se denominan a las proteínas reguladoras de operones de resistencia a tetraciclina. Estas proteínas son altamente similares en secuencia y función, presentando tan sólo pequeñas diferencias a nivel de la secuencia de los aminoácidos. De este abanico, en este trabajo se eligió la proteína reguladora TetR del transposón Tn10, dado su amplio estudio y uso en la construcción de dispositivos genéticos. Se trata del regulador transcripcional del operón de resistencia antibiótica *tet* (Figura 1-5). Se expresa del gen *tetR*, que se encuentra localizado en sentido contrario a *tetA*, gen que codifica para la proteína de resistencia (Jorgensen and Reznikoff, 1979). La forma nativa de TetR es el dímero, con una constante de equilibrio de dimerización superior a 10^7 M^{-1} (Biliouris *et al.*, 2011; Hillen *et al.*, 1983).

El homodímero de TetR presenta baja afinidad por secuencias de DNA inespecífico pero se une con gran afinidad a su operador, con una constante de asociación de $\approx 10^{11} \text{ M}^{-1}$ en condiciones fisiológicas (Lederer *et al.*, 1995). Cada operador es una secuencia palíndroma con un par de bases insertado en su eje de simetría (Wray and Reznikoff, 1983). En la región entre los genes *tetA* y *tetR* se localizan sus promotores y dos operadores para TetR, *tetO*₁ y *tetO*₂ (Bertrand *et al.*, 1983). La distancia entre los ejes de simetría de los operadores varía entre 27 y 36 en los distintos operones *tet* (Saenger *et al.*, 2000). Las constantes de unión de los operadores con el dímero de TetR son del mismo orden de magnitud, 10^2 - 10^3 M^{-1} . Según Hillen, los dos operadores tienen afinidad similar (Hillen *et al.*, 1983), *tetO*₁ es dos veces menos afín que *tetO*₂ según (Wissmann *et al.*, 1986), y 5 veces menos según (Kleinschmidt *et al.*, 1991). La unión de uno de los operadores con un dímero no modifica la afinidad del otro por el regulador, lo que indica poca o ninguna cooperatividad (Kleinschmidt *et al.*, 1988). Los promotores de la región intergénica se disponen en sentido divergente y solapados en parte (Hillen *et al.*, 1984). Existe un promotor para el gen *tetA*, que se solapa por ambos operadores y, por lo tanto, la expresión de TetA se reprime al ser ocupados cualquiera de los dos operadores. El gen *tetR*, en cambio, tiene dos promotores, *p*_{R1} y *p*_{R2}, cada uno solapado respectivamente con *tetO*₁ y *tetO*₂. TetR ve levemente impedida su expresión al ser ocupado *tetO*₂, y completamente al ser ocupado *tetO*₁ (Meier *et al.*, 1988).

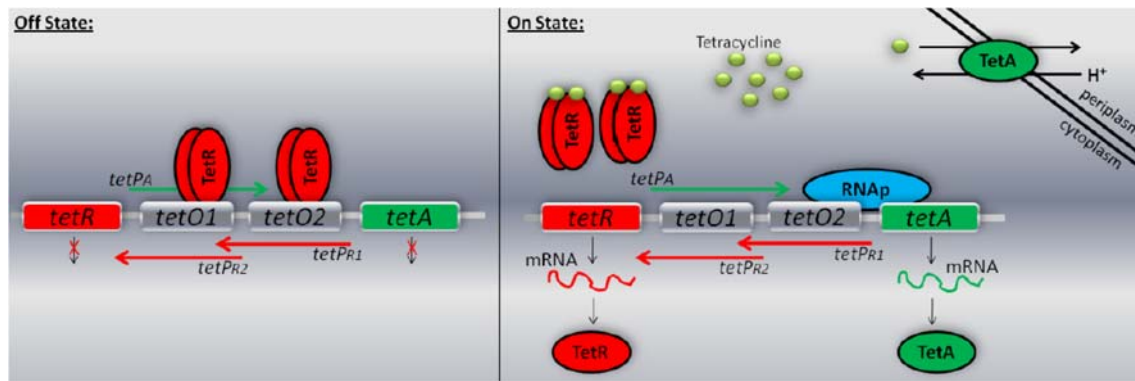


Figura 1-5. Operón *tet* y su mecanismo de regulación. Cuando no hay tetraciclina, las células ahorran energía reprimiendo la expresión de *tetA* y *tetR* (*Off State*, panel de la izquierda). Cuando hay tetraciclina, se produce TetA, que bombea Tc fuera de las células y se produce más TetR para apagar nuevamente la expresión de *tetA* cuando ya no sea necesaria (*On State*, panel de la derecha). Tomado de (Biliouris *et al.*, 2011).

Como sistema de resistencia, el operón *tet* responde a la presencia de tetraciclina. Tras difundir por la membrana, la Tc se combina con diferentes cationes, especialmente Mg^{2+} . Esta combinación es la responsable de la actividad antibiótica. Se une a la subunidad 30S del ribosoma, con una constante de asociación de $10^6 M^{-1}$ (Epe and Woolley, 1984), e impide la síntesis proteica. TetR se une a dos complejos $[Mg-Tc]^+$ con una K de $10^9 M^{-1}$ (Takahashi *et al.*, 1986). La unión provoca un cambio conformacional en el dímero de TetR que impide su unión con el operador, liberándolo. La unión de la primera molécula de Tc produce un descenso de 10^2 - 10^3 veces de la afinidad de TetR por su operador, mientras que la unión de la segunda molécula de Tc a la otra subunidad del dímero de TetR conlleva un descenso de otras 10^4 - 10^7 veces (Lederer *et al.*, 1995).

La liberación de los promotores da lugar a la expresión de los dos genes. TetA es una proteína integral de membrana que se encarga de la exportación del antibiótico de la célula, pero su sobreexpresión provoca pérdida del potencial de membrana y muerte celular (Eckert and Beck, 1989). La expresión conjunta de TetR permite mantener controlada la producción de TetA y evita, por tanto, tales consecuencias.

1.4. Regulación analógica frente a regulación digital

Como ilustra el caso de la regulación por TetR, las bacterias utilizan reguladores transcripcionales para modular la síntesis de proteínas estructurales y metabólicas en respuesta a distintas señales. Sin embargo, aunque conocemos con bastante detalle los mecanismos moleculares que median en la regulación transcripcional, nuestro conocimiento de la relación cuantitativa entre señal y respuesta sigue siendo muy limitado. La cuestión fundamental radica en determinar cómo responde la célula ante la presencia de la señal. ¿Son los TFs capaces de transmitir información acerca de la magnitud de la señal o sirven

simplemente para identificar su presencia? Esta pregunta remite a las dos formas fundamentales de controladores que se utilizan en circuitos electrónicos. Los controladores analógicos elaboran respuestas de acuerdo a la intensidad de la señal recibida. Son capaces, por tanto, de interpretar y responder a cambios de magnitud, de forma que un incremento en la señal se traduce en un cambio proporcional en la respuesta. En el otro extremo, los controladores digitales se basan en respuestas de todo o nada. Un controlador digital actúa cuando la señal supera un determinado valor umbral, pero incrementos de la señal a partir de ese umbral no se traducen en cambios en la respuesta. En el caso de la regulación transcripcional, la pregunta es cómo responde la célula a variaciones en la concentración de regulador. Si los reguladores transcripcionales actúan de forma analógica, un cambio en la concentración de un represor, como TetR, se traducirá en un descenso proporcional en la producción de TetA. Si, por el contrario, TetR actúa como un controlador digital, solamente aquellos cambios en la concentración de TetR que impliquen cruzar el valor umbral producirán cambios en la producción de TetA. A la relación cuantitativa entre la señal (TetR) y su respuesta (TetA) la denominamos función de regulación genética o función de transferencia.

1.5. Funciones de transferencia

Nuestro conocimiento acerca de las relaciones cuantitativas entre la concentración de un TF y la respuesta de su promotor diana provienen, fundamentalmente, de estudios *in vitro* y modelos basados en cinética química. El modelo general asume que el TF experimenta difusión en tres dimensiones, de tal forma que la tasa de asociación a la secuencia operadora sigue una cinética de orden 1 respecto a la concentración del TF. Se asume también que la disociación del TF de su sitio diana sigue una cinética de orden 0 respecto a la concentración de TF (para cada complejo TF-Promotor, existe una probabilidad no nula constante de que el TF se suelte de su operador). Bajo estas asunciones, y en condiciones de equilibrio, la proporción de promotores libres de TF (que se toma como equivalente al tiempo promedio experimentado por el promotor en estado libre) se relaciona con la concentración de TF mediante una ecuación de tipo Michaelis-Menten:

$$\frac{P}{P_t} = \frac{K}{K + X}$$

Ecuación 1-1. Michaelis-Menten.

Donde P/P_t es la proporción de promotores libres de TF, K es la constante de disociación y X la concentración del TF. Cuando el TF se une a más de un operador en el mismo promotor, la ecuación responde a una curva de tipo Hill, en la que aparece un exponente n correspondiente

a la no-linealidad en la respuesta. Este valor n proviene de aproximar el producto de constantes como una única constante elevada a un exponente. Es por esto que n corresponde al número de sitios de unión siempre y cuando todos estos sitios presenten una misma K . Para el resto de situaciones, n es solamente un valor aproximado y puede tomar valores no enteros. Tanto K como n son específicas de un par TF/promotor regulado y, en principio, dependen exclusivamente del mecanismo de unión entre la proteína y su sitio diana. Para convertir estas funciones de unión en funciones de transferencia, se considera que la producción del *output* (en este caso producción de mRNA) es proporcional a la ocupación del promotor. De este modo, el cambio en la concentración de un mRNA seguiría:

$$\frac{d[mRNA]}{dt} = \underbrace{\lambda \frac{P}{P_t}}_{\text{Síntesis}} - \underbrace{\beta[mRNA]}_{\text{Degradación}}$$

Ecuación 1-2

Donde λ indica la tasa de mRNA por unidad de tiempo producida cuando todo el promotor está libre de unión a TF, mientras que β indica la tasa de degradación del mRNA. Sustituimos ahora la fracción de promotor libre por las ecuaciones correspondientes al equilibrio de unión y, en el caso de un mRNA controlado por un represor transcripcional, se obtiene:

$$\frac{d[mRNA]}{dt} = \underbrace{\lambda \frac{K^n}{K^n + X^n}}_{\text{Síntesis}} - \underbrace{\beta[mRNA]}_{\text{Degradación}}$$

Ecuación 1-3

A esta ecuación, que correlaciona la concentración del TF con la síntesis de su gen diana, la denominamos función de transferencia o función de regulación genética. Funciones de transferencia de tipo Hill se infieren no solo a través de la cinética química, sino en base a la termodinámica o estudios probabilísticos. De hecho la función de transferencia adopta una forma de tipo Hill debido a dos asunciones implícitas en todos estos modelos:

- i) Que la difusión del TF se realiza en 3D, de tal manera que la probabilidad de encontrar el sitio de unión es independiente de la posición del TF en cada momento.
- ii) Que la escala temporal de producción/degradación del mRNA es mucho mayor que la escala temporal de unión/separación del TF a su sitio diana, de tal manera que podemos asumir un equilibrio de unión.

Una aproximación alternativa consiste en eliminar la aproximación de Hill y considerar el producto de diferentes constantes de manera independiente (Carbonell-Ballester *et al.*,

2014). Esta aproximación tiene la ventaja de que permite considerar de manera separada las distintas etapas del proceso (dimerización del TF, unión al promotor, traducción del mRNA, etc.). Sin embargo, mantiene las asunciones fundamentales de las aproximaciones anteriores: difusión en 3D y equilibrio rápido de unión en estado estacionario.

1.5.1. Métodos experimentales para medir la función de transferencia

La medida de la función de transferencia consiste en observar la respuesta del sistema (la proteína regulada) a distintas concentraciones del *input* (TF). Aunque el proceso es conceptualmente sencillo, su ejecución experimental plantea problemas. Producir un rango suficiente de concentraciones del *input* es sencillo cuando lo que se intenta medir es la función que correlaciona un inductor químico con la respuesta de un promotor regulado por un TF que responde a ese inductor (IPTG para LacI, Tc para TetR, etc.). Sin embargo, la mayor parte de TFs carecen de inductores químicos; y para aquellos que sí los tienen, la función de transferencia respecto al inductor y al TF no es necesariamente la misma. Para producir un rango suficiente y mensurable de TF, Rosenfeld y colaboradores elaboraron una aproximación experimental basada en la dilución progresiva del TF por el crecimiento celular (Rosenfeld *et al.*, 2005). El método de Rosenfeld se basaba en el bloqueo de la síntesis del TF, mientras que se permitía a las células seguir creciendo (Figura 1-6.A). Al multiplicarse las células diluían de manera continua la concentración de TF, de forma que midiendo la respuesta del promotor diana se podía obtener una estimación de su dependencia frente a la concentración de TF. El método de Rosenfeld presenta, no obstante, un problema importante. Dado que la concentración de TF es transitoria (las células se encuentran creciendo de manera continua), no es posible realizar mediciones del *output* en estado estacionario. Esto hace que la respuesta del *output* deba medirse en base a la derivada de la concentración, medidas altamente ruidosas desde el punto de vista experimental. Para conseguir medir en estado estacionario y no tener que recurrir a la medición de derivadas, nuestro grupo planteó un segundo método basado en la expresión controlada del TF (Fernandez-Lopez *et al.*, 2010). Este método se basa en obtener inducciones graduales utilizando el promotor de la arabinosa p_{BAD} en una cepa que expresa constitutivamente los transportadores de arabinosa (*Escherichia coli* BW27783). Esto permite mantener la concentración del TF a distintos niveles de manera estable, pudiéndose medir la respuesta del promotor regulado en estado estacionario (Figura 1-6.B).

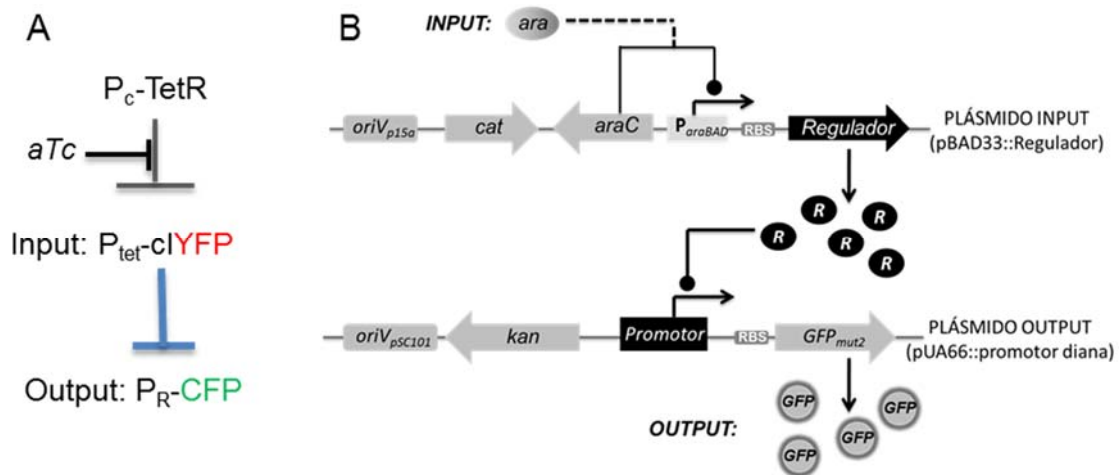


Figura 1-6. (A) Esquema del método de Rosenfeld. El represor transcripcional cl es marcado mediante fusión traduccional con YFP y se expresa desde p_{tet} . TetR constitutivamente producido bloquea su expresión. Añadiendo aTc puede liberarse y la expresión del promotor diana p_R es reportada por la fluorescencia de CFP. (B) Método de Fernández-López. Mediante inducción por arabinosa se regula la cantidad de regulador producido desde el vector pBAD33. Este regulador controla la expresión del promotor diana cuya actividad es mostrada por la fluorescencia de la GFP en el vector pUA66. Tomado de (Fernandez-Lopez *et al.*, 2010).

1.5.2. Estocasticidad en la expresión genética y sus consecuencias en la función de transferencia

La función de transferencia entre un TF y su promotor regulado no depende exclusivamente de las características bioquímicas de la unión del TF a sus operadores. La conversión del *input* (concentración de TF) a un *output* (concentración de proteína regulada) requiere de la transcripción y traducción del gen regulado. Es por esto que la estocasticidad en los procesos de síntesis y degradación del mRNA y las proteínas tienen un impacto directo sobre la naturaleza de la función de transferencia. La naturaleza estocástica de las reacciones químicas tiene importantes consecuencias en los procesos de expresión genética (Eldar and Elowitz, 2010; Elowitz *et al.*, 2002), debido a tres propiedades fundamentales de dichos procesos:

- i) A diferencia de lo que sucede con metabolitos y otras biomoléculas, el número de mRNAs y proteínas por célula suele ser bajo (por ejemplo, el propio DNA cromosómico está en 1-4 copias por célula). Dado que la magnitud relativa de las fluctuaciones es mayor cuanto menor es el número promedio de moléculas, aquellas especies moleculares presentes en bajo número sufren de mayores niveles de ruido en su expresión (Eldar and Elowitz, 2010; Elowitz *et al.*, 2002; Paulsson, 2004; Sanchez and Golding, 2013).
- ii) Traducción y transcripción son procesos conectados. Incluso aunque se produjeran miles de proteínas a partir de un mRNA, si este mRNA está presente en un bajo número de copias, tanto el mRNA como la proteína sufrirían importantes fluctuaciones (Pedraza

and van Oudenaarden, 2005).

- iii) Además de las fluctuaciones producidas por sus propios procesos de síntesis y degradación (que se denominan genéricamente ruido intrínseco), mRNA y proteínas están sujetos también a las fluctuaciones en la maquinaria general de la célula (RNA polimerasa, RNAsas, proteasas, ribosomas, etc.). A esta fuente de fluctuaciones se la denomina ruido extrínseco (Elowitz *et al.*, 2002; Swain *et al.*, 2002).

Las fluctuaciones en la expresión genética implican un límite a la capacidad celular de funcionar de un modo determinista. Células genéticamente idénticas situadas en el mismo ambiente darán lugar a niveles de expresión diferentes. La expresión de un gen, por tanto, no puede ser descrita con un número sino que corresponde a una distribución estadística, con una cierta media y una desviación estándar asociada (Paulsson, 2004). Un análisis más detallado de los formalismos matemáticos utilizados para describir la expresión genética estocástica se da en Materiales y Métodos (apartado 2.9).

La estocasticidad en la expresión genética tiene importantes consecuencias en la fisiología celular. En organismos unicelulares, permite a células genéticamente idénticas diferenciarse de manera espontánea sin necesidad de señales ambientales, lo que tiene importantes consecuencias evolutivas (Norman *et al.*, 2015), pero también plantea problemas en el control de los procesos celulares (Lestas *et al.*, 2010). Dado que señales y componentes sufren fluctuaciones, la célula funciona en un estado de indeterminación permanente (Bowsher and Swain, 2014; Lestas *et al.*, 2010). En el caso concreto de la función de transferencia, la estocasticidad en la expresión genética se traduce en que, para una concentración dada de *input* (TF) es esperable obtener una distribución de *outputs* distintos. De igual manera, un *output* determinado puede ser producido por varios *inputs* distintos. El ruido en la expresión genética hace que la relación TF/proteína regulada no sea unívoca, sino estadística. Además de disminuir la capacidad de control de la célula, pasar de una relación unívoca a una relación estadística tiene consecuencias en cómo se comportan las células individuales frente a la población en su conjunto.

En una relación determinista, en la que la relación *input-output* es unívoca, la función de transferencia de cada célula individual es equivalente a la observada para la población en su conjunto. Esto es así porque, si la regulación genética funcionase de modo determinista, las concentraciones totales de *input* y de *output* observadas para N células no serían más que N veces la concentración individual de cada célula. Podríamos pensar que, en el caso de que existan fluctuaciones en torno a la media, el valor individual de cada célula podría aproximarse

al valor poblacional promedio simplemente teniendo en cuenta la desviación estándar de la población. Desgraciadamente, en el caso de la función de transferencia, esto sólo es así si la función es lineal. La Desigualdad de Jensen (DJ), establece que si aplicamos una función matemática a una distribución dada, la media de la función (en nuestro caso, el promedio de *outputs*) es igual a la función promedio (en nuestro caso, el resultado de aplicar la función al promedio de *inputs*), si y solo si la función es lineal (Jensen, 1906). La razón de la DJ radica en que funciones altamente no lineales no transforman de manera proporcional los extremos inferiores y superiores de una distribución. Tomemos por ejemplo una distribución con promedio 10 y apliquemos una función de transferencia consistente en elevarla a la cuarta potencia. Con una desviación en el *input* de nueve unidades tenemos $f(x) = 1^4 = 1$ en un extremo, y $f(x) = 19^4 = 130321$ en el otro. La DJ asegura, además, que si la función es convexa entonces $f(\langle x \rangle) \leq \langle f(x) \rangle$. Donde los valores entre $\langle \rangle$ indican el valor promedio. Para el caso concreto de la función de transferencia en la regulación genética, la DJ implica que las funciones obtenidas a partir de promedios poblacionales podrían ser poco o nada representativas de la función de transferencia intracelular. En concreto, una función de transferencia medida a partir de valores promedios estaría sesgada por los valores más altos de la distribución. Dado que las respuestas las elaboran las células individuales, no la población en conjunto, se hace necesario comparar las funciones de transferencia obtenidas a nivel poblacional con las respuestas individuales de las células dentro de la población.

1.5.3. Función de transferencia y toma de decisiones

Las discrepancias entre los promedios poblacionales y las células individuales subrayan la conexión entre las funciones de transferencia transcripcionales y los procesos de toma de decisiones en la célula. Como apuntamos anteriormente, la regulación transcripcional constituye el mecanismo fundamental que las bacterias utilizan para elaborar respuestas frente a señales y estímulos ambientales. Dado que señales y componentes están asociados a unas fluctuaciones, tanto la detección como la respuesta a una señal dada esta sometida a un cierto grado de incertidumbre. Por esta razón, la consecuencia fenotípica de la función de transferencia es equivalente a una toma de decisión por parte de la célula.

¿Qué estrategia deben adoptar las células para interpretar y hacer uso de las señales ruidosas? Las estrategias de procesamiento de información utilizadas por las células para la toma de decisiones son probabilísticas y ocurren en tres niveles (Perkins and Swain, 2009) (Figura 1-7): i) las células deben inferir las señales probables diferenciándolas de las del ruido, anticipándose al estado futuro del ambiente en el que se encuentran; ii) deben sopesar los costes y

beneficios de cada respuesta potencial, teniendo en cuenta ese futuro probable; iii) la decisión es tomada en presencia de competidores potenciales, que también toman decisiones. Perkins y Swain han propuesto que las estrategias de toma de decisión están conservadas, con un número reducido de estrategias implementadas en mecanismos bioquímicos diferentes en muchos organismos (Perkins and Swain, 2009), lo que ligaría la Biología de Sistemas con la Biología Evolutiva.

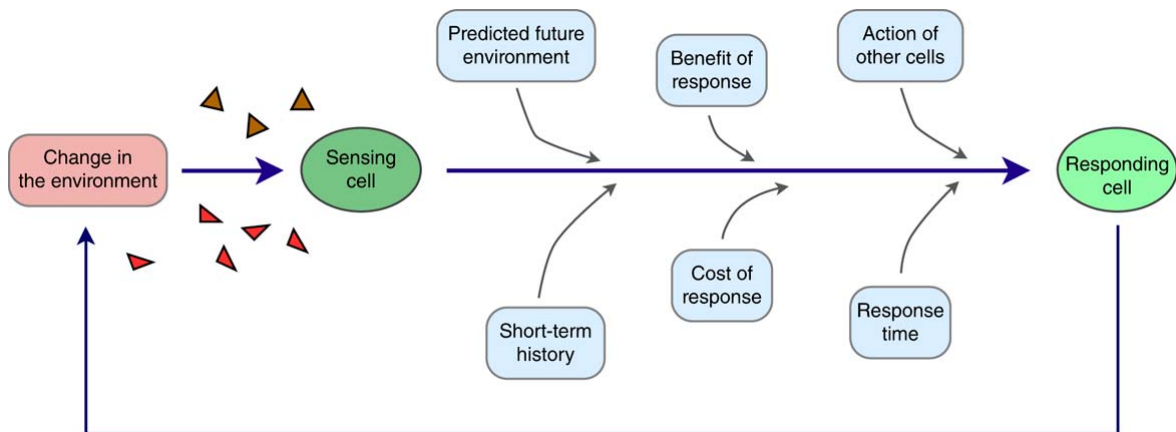


Figura 1-7. Factores que influyen en la toma de decisión por la célula. Una célula siente señales generadas por un cambio en el ambiente y debe decidir una respuesta adecuada. Esta toma de decisión depende de las predicciones celulares acerca del estado presente y futuro de su ambiente basado en las señales que ha recibido; el historial reciente de la célula y los costes y beneficios esperados para cada respuesta potencial; las acciones de otras células, que pueden ser competidores directos o cooperadores; así como en el tiempo requerido para tomar la decisión y generar la respuesta, que puede ser un cambio en el estado interno o una acción que modifique el ambiente. Tomada de (Perkins and Swain, 2009).

1.6. Entropía de una distribución

El hecho de que *input* y *output* en una función de transferencia aparezcan como distribuciones probabilistas implica de manera inmediata que es posible asociar a ambas variables un cierto grado de incertidumbre. Shannon se preguntó si era posible dar a esa incertidumbre un valor cuantitativo. De manera análoga, si dos variables aleatorias están relacionadas de alguna forma, la observación de una de ellas debe llevar asociada una reducción en la entropía de la otra. En su obra seminal (Shannon, 1948), Shannon razonó que esta disminución cuantitativa es una medida de la información que una variable aporta sobre la otra, y que para que esta medida sea adecuada, debe cumplir ciertos requisitos:

- i) Si cada una de las N realizaciones (valores individuales) posibles de la distribución son equiprobables, entonces la información transmitida al observar un valor particular debe crecer con N .
- ii) Si la inferencia puede ser descompuesta en varias etapas, entonces la información transmitida en cada una de las etapas debe contribuir a la información global.

- iii) Si la inferencia puede ser dividida en un árbol de decisiones, entonces la información global debe ser una suma ponderada de las informaciones parciales a lo largo de una trayectoria en el árbol de decisiones.

Shannon demostró que la única métrica capaz de satisfacer estas condiciones es la entropía de la distribución $H(p)$:

$$H(p) = -k \sum_N p_N \log(p_N)$$

Ecuación 1-4. Entropía de Shannon.

Donde p_N es la probabilidad de cada una de las N realizaciones y k es una constante arbitraria. Esta cuestión es equivalente a preguntarse cuántas unidades de la constante k necesitamos para determinar de manera precisa la incertidumbre asociada a la distribución. Esto a su vez depende del valor de k , la unidad de incertidumbre, que establece el "alfabeto" en el que expresamos la entropía. Cuando $k=1$, estamos indicando que la entropía está expresada en un árbol binario de decisiones con probabilidad básica $1/2$ (el número de respuestas sí/no necesarias para llegar al valor de la realización entre los N posibles). En ese caso, los logaritmos se toman en base 2 y a esa unidad de entropía la denominamos bit. Es importante resaltar que la elección de la base únicamente cambia la unidad en la que se expresa la entropía: para base 2 la unidad es el bit, para base e la unidad es el nat y cuando la probabilidad básica es $1/10$ la unidad se denomina hart. De esta forma, es posible asignar un valor numérico a la incertidumbre asociada a una distribución dada.

1.7. Información mutua

Dado que es posible asignar un valor numérico a la incertidumbre de una distribución, debería ser posible también asignar un valor numérico a la reducción en la incertidumbre sobre una variable cuando observamos una variable relacionada. Denominamos información mutua entre dos variables X e Y , $I(X; Y)$, a la disminución en la entropía de una de estas variables cuando observamos la otra, de esta forma:

$$I(X; Y) = H(X) - H(X|Y)$$

Ecuación 1-5

Donde $H(X|Y)$ es la entropía condicional: la entropía de la distribución en X dado un valor de Y . Por las propiedades de la entropía condicional, es posible reordenar la ecuación anterior y obtener la igualdad:

$$I(X; Y) = H(X) + H(Y) - H(X, Y)$$

Ecuación 1-6

Esto es, la información mutua entre dos variables aleatorias es igual a la suma de las entropías de sus distribuciones marginales menos la entropía de la distribución conjunta. La información mutua posee una serie de cualidades importantes:

- i) Como su propio nombre indica, es mutua, esto es $I(X; Y) = I(Y; X)$.
- ii) Cuando dos variables son independientes entonces $H(X, Y) = H(X) + H(Y)$, lo que implica $I(X; Y) = 0$.
- iii) Al ser una suma de probabilidades, $I(X; Y) \geq 0$.
- iv) Es invariante para transformaciones no lineales sobre X e Y . Esta propiedad resulta extraordinariamente útil a la hora de medir experimentalmente $I(X; Y)$, dado que este valor puede inferirse de las distribuciones reales (número de moléculas, por ejemplo) como de alguna función de esta distribución (fluorescencia de una GFP, por ejemplo) incluso cuando la relación fluorescencia/número de moléculas es altamente no lineal.

Además de estas propiedades la información mutua tiene la ventaja de que puede ser interpretada en función del número de estados que la célula es capaz de discriminar (Bialek, 2012). Así, si $I(X; Y) \leq 1$ bit, la célula solo será capaz de discriminar la señal de manera binaria (ausencia/presencia de la señal) de forma que cuanto menor sea $I(X; Y)$ más ruidosa será la inferencia. Valores superiores de I indican que el sistema es capaz de cuantificar la señal, y cuanto mayor sea el número de bits, mayor será el número de estados que el sistema es capaz de resolver de manera no ambigua. Esto no quiere decir, no obstante, que el número de bits indique exactamente el número de estados posibles de la célula: un sensor con $I=1,6$ bits no indica necesariamente que el número de estados sea 3, si no que puede corresponder a un estado activado, un estado inactivado y unos valores intermedios que son inferidos con alto nivel de ruido. En otras palabras, la información mutua cuantifica la capacidad del canal, pero no indica necesariamente la forma ni el número de estados de este.

1.7.1. Uso de la información mutua en la regulación transcripcional

La mayor parte de las aplicaciones de la Teoría de la Información en Biología se han circunscrito al análisis de la comunicación neuronal. Esto es debido a que la comunicación por impulsos eléctricos en las neuronas es similar al canal de información que Shannon tenía en

mente cuando concibió las bases de la Teoría. En concreto, tanto un axón neuronal como un cable de comunicaciones es un canal que permite la transmisión de múltiples impulsos a lo largo del tiempo. Son canales rápidos que permiten la transmisión casi instantánea de información. En cambio, en un circuito transcripcional la información se transmite mediante la unión o separación del TF a sus sitios en el DNA, reacciones mucho más lentas. Aunque esto limita la aplicación de algunos de los teoremas más importantes de la Teoría de la Información (en concreto, el Segundo Teorema de Shannon o de codificación en un canal ruidoso), no prescribe que la información mutua sea una herramienta poderosa a la hora de analizar los circuitos transcripcionales y las consecuencias del ruido en la expresión genética en la toma de decisiones celulares. La aplicación de la información mutua a reguladores transcripcionales se debe a Tkačik y Bialek, que en una serie de trabajos demostraron la capacidad informativa del factor de regulación Hunchback en la morfogénesis de la larva de la mosca del vinagre *Drosophila melanogaster* (Gregor *et al.*, 2007a; Gregor *et al.*, 2007b; Tkacik *et al.*, 2008). Durante el desarrollo embrionario de la larva de la mosca, ciertos factores de transcripción se secretan para formar gradientes de un polo a otro del embrión. Estos gradientes indican la posición relativa de unas células respecto de otras, de tal forma que dependiendo de la dosis de TF que cada célula reciba se diferenciará en una parte distinta de la mosca adulta. Hunchback es un determinante del eje anteroposterior. En el embrión temprano existe un gradiente desde un polo al otro de la larva que indica a las células si van a formar parte de la cabeza o del tórax de la mosca. Bialek y colaboradores aprovecharon que el propio embrión genera un gradiente completo (la distribución marginal de Hunchback) y midieron la distribución de uno de sus genes diana. A partir de estas distribuciones pudieron inferir que la información transmitida correspondía a casi 2 bits de información, por lo que Hunchback no actuaba delimitando únicamente dos regiones, si no que transmitía información posicional más precisa. De manera similar, Selimkhanov y colaboradores (Selimkhanov *et al.*, 2014) estudiaron la respuesta de los reguladores Ca, ERK y NFkB en células de embriones de la rana *Xenopus laevis*, encontrando en todos los casos $I < 1$ bit. Es importante apuntar que en organismos eucariotas, la regulación transcripcional presenta un paso adicional que es el de la migración del TF al núcleo. En muchos casos el inductor produce un cambio conformacional que permite el transporte activo del TF al compartimento nuclear, donde ejerce su acción. Esta dinámica de transporte está detrás de una característica importante de la regulación transcripcional en eucariotas. A diferencia de lo que se ha descrito en bacterias, los organismos eucariotas parecen capaces de controlar la expresión de genes en función tanto de la amplitud de la concentración de un TF como de la frecuencia de pulsos de introducción al núcleo. Esta respuesta por "frecuencia modulada" ha sido observada en multitud de reguladores

eucarióticos, como p53 (Purvis *et al.*, 2012), los reguladores asociados al calcio como Crz1 (Cai *et al.*, 2008) o el factor de estrés Mns2 en levaduras (Hao and O'Shea, 2011). Este hecho ha abierto la discusión sobre si la información en los TFs eucarióticos se transmite fundamentalmente mediante inferencia de la concentración del TF o, por el contrario, la información está contenida en la dinámica temporal del TF. Mientras que Selimkhanov y colaboradores (Selimkhanov *et al.*, 2014) sostienen que la información contenida en la secuencia temporal es mayor que la aportada por la concentración, Hansen y O'Shea (Hansen and O'Shea, 2015) sostienen que, en el caso de Mns2, tanto la información modulada por amplitud como por frecuencia es similar, y de carácter binario.

Pese a que los organismos procarióticos presentan mecanismos transcripcionales más sencillos que los eucariotas, no existen análisis experimentales sobre la cantidad de información que las funciones de transferencia transcripcionales son capaces de transmitir en bacterias. La razón fundamental radica en la dificultad en obtener un muestreo completo de la distribución marginal del *input*. Como indicamos anteriormente, la información mutua está condicionada por la entropía de las variables analizadas. En eucariotas, la separación física entre citoplasma y núcleo permite la introducción controlada de TF mediante distintas dosis de inductor. En procariotas no existe un método similar. Los inductores procarióticos modifican la afinidad del TF, pero no es posible cuantificar en células individuales la cantidad de TF unida al inductor. Sin este parámetro, no es posible obtener la distribución completa del TF.

Objetivos

Objetivos:

1. Caracterizar la función de transferencia entre el regulador transcripcional TetR y su promotor diana p_{tetA} en la bacteria *Escherichia coli*.
2. Comparar la eficiencia y precisión de medidas poblacionales frente a medidas en células individuales en la obtención de la función de transferencia.
3. Obtener la información mutua entre el regulador TetR y su promotor diana p_{tetA} , determinando si TetR actúa como un regulador analógico o un regulador digital.
4. Estudiar los parámetros que relacionan la función de transferencia con la información transmitida por el canal transcripcional.

Materiales y Métodos

2. Materiales y Métodos

2.1. Cepas

Las cepas utilizadas en el presente trabajo se detallan en la Tabla 2-1. Todas son *E. coli* derivadas de la cepa K-12.

Tabla 2-1. Cepas utilizadas.

Nombre	Genotipo	Referencia
DH5α	<i>F⁻ endA1 hsdR17 (r_k⁻, m_k⁺) supE44 thi-1 λ⁻ recA1 gyrA96 relA1 deoR Δ(lacZYA-argF)-U169 φ80dlacZΔM15</i>	(Grant <i>et al.</i> , 1990)
BW27783	<i>lacI^q rrnB_{T14} lacZ_{WJ16} hsdR514 Δ(araBAD_{AH33}) Δ(rhaBAD_{LD78}) Δ(araFGH) φ(ΔaraEp P_{CP8}-araE)</i>	(Khlebnikov <i>et al.</i> , 2001)
TB10	<i>ilvG rfb50 rph1 nadA::Tn10 λcl857 Δ(cro-bioA)</i>	(Johnson <i>et al.</i> , 2004)
RSS247	<i>BW27783 Δ(araBAD)::(tetR::mkate2 p_{tetA}gfp)</i>	Esta tesis
RRS252	<i>BW27783 Δ(araBAD)::(tetR::mkate2 p_{tetR} p_{tetA}gfp)</i>	Esta tesis
RRS279	<i>BW27783 Δ(araBAD)::(tetR::gfp p_{tetA}mkate2)</i>	Esta tesis
RRS283	<i>BW27783 Δ(araBAD)::(p_{tetA}gfp)</i>	Esta tesis

2.2. Cultivo

El cultivo de las cepas, salvo excepciones dependientes de protocolos puntuales, se realizaron en condiciones estándar para *E. coli*. Salvo excepciones dependientes de protocolos puntuales, los cultivos líquidos se crecieron a 37 °C en agitación (120-150 rpm) en matraces de cristal, y el crecimiento en sólido se realizó en placas de Petri de 90x14 mm en cámaras de 37 °C.

2.2.1. Medios de cultivo

Para el crecimiento habitual de los cultivos se utilizó medio líquido LB de Sigma (10 g de triptona, 5 g de extracto de levadura, 5 g de NaCl por litro), o LB agar de Sigma (LB + 15g de agar por litro) para crecimiento en sólido.

Para ensayos que necesitaron de unas condiciones de crecimiento enriquecidas se utilizó medio de cultivo SOB de Amresco (20 g de triptona, 5 g de extracto de levadura, 0,5 g NaCl y 20 ml de MgSO₄ 1M por litro).

Cuando la autofluorescencia del LB suponía un inconveniente se utilizó en su lugar medio M9 de Fluka-Sigma (6g de Na₂HPO₄, 3g de KH₂PO₄, 1g de NH₄Cl y 0,5g de NaCl por litro) suplementado con CaCl₂ (100 μM), MgSO₄ (20 mM), casaminoácidos (0,2 %), y glucosa o glicerol (0,5 %) como fuente de carbono. En el caso de crecimiento en sólido para microscopía, a la mezcla se le añadió agarosa para biología molecular de Sigma (1,5 %).

2.2.2. Antibióticos y correguladores

Las cepas se cultivaron en medio selectivo añadiendo antibióticos correspondientes para cada caso. Los antibióticos se prepararon en stocks 1000x para luego añadirlos al medio y conseguir la concentración adecuada:

Tabla 2-2. Preparación y uso de los antibióticos.

Antibiótico	Compuesto	Soluto	Stock	Concentración cultivo
Ácido nalidíxico	Ácido nalidíxico (Apollo)	Agua	20 mg/ml	20 µg/ml
Ampicilina	Ampicilina sódica (Apollo)	Agua	100 mg/ml	100 µg/ml
Cloranfenicol	Cloranfenicol (Apollo)	Etanol 100%	25 mg/ml	25 µg/ml
Kanamicina	Sulfato de kanamicina (Apollo)	Agua	50 mg/ml	50 µg/ml
Tetraciclina	Hidrocloruro de tetraciclina (Apollo)	Etanol 50%	10 mg/ml	10 µg/ml

Además, en esta tesis se han utilizado un par de compuestos como inductores o correguladores transcripcionales. En función de las necesidades del experimento se añadieron arabinosa (L+arabinosa, Sigma) y anhidrotetraciclina (hidrocloruro de anhidrotetraciclina, Sigma) a diferentes concentraciones.

2.3. Oligonucleótidos

Los oligonucleótidos fueron sintetizados por Sigma-Aldrich o por IDT en función de la logística del instituto. Para los procedimientos de clonaje (Tabla 2-3) y secuenciación (Tabla 2-4) habituales se pidieron desalados. Cuando una mayor calidad fue necesaria, como en el caso del procedimiento “QuickChange”, se pidieron purificados por HPLC.

Tabla 2-3. Oligonucleótidos utilizados para la construcción de los dispositivos genéticos.

Nombre	Secuencia nucleotídica (5'→3')
AraC_W_74	CCTGACTAATGCGTTGGTCCTCGCGCCAGCTTAAGACGCTAATCCCTAACC TTATTTGTATAGTTCATCCATGCC
AraC_Wanner	CCTGACTAATGCGTTGGTCCTCG
AraD_W_74	ATGCTTACGCAGATAGTGTATCCAGCAGCGTTTGCTGCATATCCGGTAA AAATCGCTATAATGACCCCGAAGC
AraD_Wanner	ATGCTTACGCAGATAGTGTATCCAGC
gfpfusion_dir	AATGTGAAAGTGGGTCTAGCGGTGGCGGTGGCAGTAAAGGAGAAGAACT TTTC
gfpfusion_rev	CTCTCATCCGCCAAAACAGCCAAGCTTGCATGCCTGCAGGTGCGACTCTAGT TATTTGTATAGTTCATCCATG
GFPKpnI	CATCAGGTACCTTTAAGAAGGAGATATACATATGAGTAAAGGAGAAGAAC
GFP _{PrevHindIII}	AGCCAAGCTTATTTGTATAGTTCATCCATG
isomkate_rev	CTCTCATCCGCCAAAACAGCCAAGCTTGCATGCCTGCAGGTGCGACTCTAGT

isomkatefusdir	TATCTGTGCCCCAGTTTGC AATGTGAAAAGTGGGTCTAGCGGTGGCGGTGGCGTGTCTGAGCTGATTAA GGAG
isotetRfus_dir	ATACCCGTTTTTTTGGGCTAGCGAATTCGAGCTCGGTACCCGGGGATCCTT AAGAAGGAGATATACATATGTCTAGATTAGATAAAAAG
isotetRfus_dir	ATACCCGTTTTTTTGGGCTAGCGAATTCGAGCTCGGTACCCGGGGATCCTT AAGAAGGAGATATACATATGTCTAGATTAGATAAAAAG
isotetRfusrev	CTTAATCAGCTCAGACACGCCACCGCCACCGCTAGACCCACTTTCACATTTA AG
J23101dir	TCGAGTTTACAGCTAGCTCAGTCCTAGGTATTATGCTAGCG
J23101rev	GATCCGCTAGCATAATACCTAGGACTGAGCTAGCTGTAAAC
pBADSEVA1	CGGACTAGTATAATGTGCCTGTCAAATGGA
pBADSEVA2	ACCTTAATTAATGAGCGGATACATATTTGAATGT
pCSmkate_dir	CAAACCTGGGGCACAGATAAATGTCCAGACCTGCAGGCATGCAAGCTCTAG
pCSmkate_rev	CCTTAATCAGCTCAGACACCATATGTATATCTCCTTCTTAAATCTAGAGG
pRRG69_rev	TAAATGTCCGCAAAAAACCCC
pRRG74_GFPdir	CAGTGATAGAGAAAAGTGAAATGAGTAAAGGAGAAGAAGT
pRRG74_GFPprev	TTCACACAGGAGGCCGCTAGGCCGCGGCCGCGCAATTCTTATTTGTATA GTTTCATCCATG
pRRG74_mkatedir	CTTAAATGTGAAAAGTGGGTCTAGCGGTGGCGGTGGCGTGTG
pRRG74_mkaterrev	AGGTCATTACTGGATCTATCAACAGGAGTCCAAGACTAGTTTATCTGTGCC CCAGTTTGC
pRRG74_ptetArev	AGTTCTTCTCCTTTACTCATTTCACTTTTCTCTATCACTG
pRRG74_TetRrev	GACACGCCACCGCCACCGCTAGACCCACTTTCACATTTAAG
pTetA1Xhol	AACTCGAGTCACTGATAAACATAATATGTCC
pTetA2BamHI	ATGGATCCGCTCTGGCCTCTGACCAGCC
ptetAGFPKm_revPacI	CCTTTAATTAATATCATCAACAGGCTTACCCG
ptetAmkate_dir	CCTCTAGATTTAAGAAGGAGATATACATATGGTGTCTGAGCTGATTAAAG
ptetAmkate_rev	CTAGAGCTTGCATGCCTGCAGGTCTGGACATTTATCTGTGCCCCAGTTTG
pUA66terB1007	TGCAAGCTTGCGAAAAAACCCCGCCCTGTCAGGGGCGGGGTTTTTTGCGG ACATTTATTTGTATAGTTTCATCCATGC
pUALAAtterB1007	TGCAAGCTTGCGAAAAAACCCCGCCCTGTCAGGGGCGGGGTTTTTTGCGG ACATTTAAGCTGTAAAGCGTAGTTTTTC
pUAmkatesinter	TTATCTGTGCCCCAGTTTGCTAG
quickgfp206dir	CTGTCCACACAATCTAACTTTTGAAAGATCCC
quickgfp206rev_2	GTAATGGTTGTCTGGTAAAAGGACAGGGCCATC
tetRgfpfusion_rev	AAGTTCTTCTCCTTTACTGCCACCGCCACCGCTAGACCCACTTTCACATTTA AG
TetRHindIII	AGGTCAAAGCTTGAAGTTACCATCACGAAAAAGGTTATGCTG
TetRKpnI	CATCAGGTACCTTTAAGAAGGAGATATACATATGTCTAGATTAGATAAAAAG TAAAGTG

Tabla 2-4. Oligonucleótidos utilizados para comprobación y secuenciación genética.

Nombre	Secuencia nucleotídica (5'→3')
GFPf	CACAACATTGAAGATGGAAGC
GFPprev	CAAAGCATTGAAGACCATACGCG
GFPseq	GGGACAACACCAAGTG
Knseq	GCGTGCAATCCATCCTTGTTCATCATGCG
Knseqrev	TCGACGTTGTCACTGAAGCGGGAAGGGAC
pBAD1	CTGTTTCTCCATACCCGTT

pBAD2	CTCATCCGCCAAAACAG
pBADseq	CGTCACACTTTGCTATGCCA
pBADseq_rev	TGGCATAGCAAAGTGTGACG
pZE05	CCAGCTGGCAATTCCGA
sevaseq1	GAGTTCTGAGGTCATTACTGGATC
sevaseq2	GGAGAGCGTTCACCGACAAA
TetRseqrev	CACCTACTACTGATAGTATG

2.4. Plásmidos

Los plásmidos utilizados en esta tesis se especifican en la Tabla 2-5, con sus características fenotípicas y una somera descripción de cada uno.

Tabla 2-5. Plásmidos utilizados en esta tesis.

Nombre	OriV	Resistencia	Descripción	Referencia
pBAD33	p15A	Cm	Vector de expresión inducible por arabinosa	(Guzman <i>et al.</i> , 1995)
pRAF22	pBR322	Cm	Vector con la proteína mKate2	Fernández-López, sin publicar
pRRG1	p15A	Cm	pBAD33 con la secuencia codificante de TetR clonada	Esta tesis
pRRG8	RK2	Amp	pSEVA121 con la región de expresión del pBAD33 clonada, expresando la proteína GFPmut2	Esta tesis
pRRG9	RK2	Amp	pSEVA121 con la región de expresión del pBAD33 clonada, expresando la proteína TetR	Esta tesis
pRRG13	SC101	Km	Vector reportero de la actividad del promotor p_{tetA}	Esta tesis
pRRG54	p15A	Cm	pBAD33 expresando la fusión traduccional TetR-mKate2	Esta tesis
pRRG55	p15A	Cm	pBAD33 expresando la fusión traduccional TetR-GFPmut2	Esta tesis
pRRG56	SC101	Km	Vector reportero con la proteína mKate2	Esta tesis
pRRG57	RK2	Amp	pRRG8 expresando la proteína GFPmut2 con la misma región 5'UTR que en el caso de TetR	Esta tesis
pRRG58	p15A	Cm	pBAD33 expresando la proteína GFPmut2 con la misma región 5'UTR que en el caso de TetR	Esta tesis
pRRG62	RK2	Amp	pRRG8 expresando la fusión traduccional TetR-mKate2	Esta tesis
pRRG63	RK2	Amp Km	Vector con la región de expresión del plásmido pBAD33 y la reportera del pUA66. Ambas regiones están separadas por el terminador BBa_B1007. Expresa TetR-mKate2 e	Esta tesis

			indica la actividad del promotor p_{tetA} mediante GFPmut2	
pRRG65	SC101	Km	Vector reportero de la actividad del promotor BBa_J23101	Esta tesis
pRRG68	RK2	Amp	pRRG8 expresando la fusión traduccional TetR-GFPmut2	Esta tesis
pRRG69	RK2	Amp Km	Vector con la región de expresión del plásmido pBAD33 y la reportera del pRRG56. Ambas regiones están separadas por el terminador BBa_B1007. Expresa TetR:GFPmut2 e indica la actividad del promotor p_{tetA} mediante mKate2.	Esta tesis
pRRG74	RK2	Amp	Vector que expresa TetR:mKate2 e indica la actividad del promotor p_{tetA} mediante GFP con la misma configuración que en el sistema natural del transposón Tn10	Esta tesis
pRRG80	RK2	Amp	pRRG8 expresando la fusión traduccional TetR-GFPmut2, con la mutación A206K	Esta tesis
pRRG81	RK2	Amp Km	pRRG69 con el mutante monomérico de GFPmut2	Esta tesis
pRRG85	RK2	Amp	pRRG63 sin la secuencia de TetR:mKate2. Control de la actividad promotora de p_{tetA} .	Esta tesis
pSEVA121	Rk2	Amp	Vector de clonaje estándar europeo	(Silva-Rocha <i>et al.</i> , 2013)
pUA66	SC101	Km	Vector reportero con la proteína GFPmut2	(Zaslaver <i>et al.</i> , 2006)
R100	IncFII	Cm, Sm, sulfonamida, Tc	Plásmido natural, contiene el transposón Tn10 con el operón de resistencia a tetraciclina.	(Sugino and Hirota, 1962)

2.5. Métodos de clonaje

2.5.1. Extracción y purificación de DNA

Para la extracción y purificación de DNA se utilizaron varios kits comerciales, dependiendo de la procedencia del mismo y de las necesidades del experimento.

-DNA plasmídico de cultivo: Para la obtención de DNA plasmídico de una cepa se utilizó “GeneJET plasmid miniprep kit” de Fermentas, siguiendo el protocolo correspondiente proporcionado por el fabricante. Este tipo de extracción es un paso necesario en la construcción de clones y secuenciación de esta tesis.

- DNA total celular: Durante el desarrollo de esta tesis fue necesario únicamente para la comprobación de recombinantes cromosómicos, siendo necesaria poca cantidad. Por ello se utilizó la matriz “Instagene” de Bio-Rad, la cual proporciona DNA sin residuos de lisis celular mediante un protocolo sencillo.

- Purificación de DNA de un gel agarosa: Para purificar el DNA tras cortar una banda de un gel de electroforesis se utilizó “GeneJET gel extraction kit” de Fermentas, siguiendo el protocolo del fabricante. Esta extracción se usó en la construcción de clones y en la preparación de muestras para secuenciación.

-Purificación de DNA de una reacción enzimática: Se utilizó “GeneJET PCR purification kit” de Fermentas, siguiendo el protocolo del fabricante. Esta extracción se usó en la construcción de clones.

2.5.2.Cuantificación del DNA

La medición de la concentración de DNA obtenida en las extracciones se calculó a partir de medidas de absorbancia a 260 nm utilizando el espectrofotómetro “NanoDrop 2000” de Thermo Scientific. Siempre se comprobó la pureza del DNA extraído teniendo en cuenta la relación 260/280 nm ($\approx 1,8$) y 260/230 nm (≈ 2) ya que puede condicionar reacciones posteriores.

2.5.3.Reacción en cadena de la polimerasa (PCR)

Esta técnica de amplificación de moléculas de DNA se empleó con varios propósitos en esta tesis, en función de los cuales la metodología y los materiales variaron.

En la amplificación de moléculas de DNA destinadas al clonaje es de capital importancia que la copia de dichas moléculas sea lo más fiel posible. Por ello, se utilizaron polimerasas comerciales de alta fidelidad. En general, se utilizaron la DNA polimerasa “Vent” de BioLabs, para moléculas de longitudes inferiores a 4kb, y la “Phusion” de Thermo, para longitudes superiores. En ambos casos se siguió estrictamente el protocolo proporcionado por el fabricante.

Las PCRs empleadas para la comprobación de clones se realizaron con una polimerasa obtenida en nuestro laboratorio. La mezcla de reacción con esta polimerasa fue: 5 μ l de tampón para PCR 10x (500 mM KCl, 100 mM TrisHCl, 1 % Triton, pH 9), 2 μ l de $MgCl_2$ 50 mM, 1 μ l de mezcla de nucleótidos 10 mM total, 0,5 μ l de cada cebador 100 μ M, 0,5 μ l de polimerasa, la colonia bacteriana o el DNA a analizar y agua “Milli-Q” hasta un volumen final de 50 μ l.

Tres termocicladores fueron utilizados durante este trabajo: “UNO II” de Biometra, “C1000 Thermal Cycler” de BioRad y “TProfessional Thermocycler” de Biometra. En general, el programa utilizado, con modificaciones en función de la polimerasa usada y la molécula a amplificar, fue:

- i) Un paso inicial de desnaturalización de 5 minutos a 95 °C,
- ii) 30 repeticiones de 3 pasos: 30 segundos a 95 °C, 30 segundos a la temperatura de hibridación DNA/cebador y 1 minuto por kb de DNA amplificado a 72 °C.
- iii) Un último paso de elongación 10 minutos a 72 °C.
- iv) Mantenimiento de la reacción a 4 °C.

2.5.4.Electroforesis de DNA en gel de agarosa

Esta técnica para separar fragmentos de distintos tamaños de DNA se empleó con diferentes objetivos, tanto para analizar resultados de PCRs, como para obtener fragmentos de DNA para su uso en clonación tras el corte de la banda adecuada del gel. Dependiendo de las longitudes de las moléculas a discernir se emplearon geles de agarosa al 1 o 2%. La agarosa se disolvió en tampón TBE 0.5x (Tris-HCl 45 mM, ácido bórico 45 mM, EDTA 0,5 mM pH 8,2). Tras hervir y disolver la mezcla correctamente y atemperar la solución hasta 60-70 °C, se añadió “SafeView Classic” de ABM, en una proporción de 5 µl por cada 100ml de agarosa. Las muestras de DNA a cargar en el gel se mezclaron en una proporción 5:1 con el tampón de carga 6x (Azul de bromofenol 0,25 % (p/v), glicerol 30 % (v/v) en tampón TBE 0,5x). En función de los tamaños moleculares objeto de análisis se emplearon distintos marcadores moleculares: “HyperLadder I” (BioLabs) para muestras de hasta 10 kb, y “Generuler 100bp DNA ladder plus” (Fermentas) para tamaños entre 100 y 3000 bp. La electroforesis se realizó en cubetas horizontales de BioRad sumergiendo el gel en tampón TBE 0,5x, y aplicando voltajes entre 100-120 V. Los geles se analizaron en un transiluminador de luz ultravioleta “Gel Doc 2000” (BioRad) y se analizaron mediante el programa “Quantity One” (BioRad).

2.5.5.Digestión de DNA mediante enzimas de restricción

Las enzimas de digestión empleadas en esta tesis pertenecen a la marca comercial Fermentas. Se aplicó el protocolo indicado por el fabricante, que en general se trata de una reacción con un volumen de 20 µl donde se mezclan 2 µl del tampón de reacción 10x, 1 µl de enzima, 500 ng de DNA y agua “Milli-Q” hasta completar. La incubación se realizó a 37 °C durante una hora, salvo para las enzimas “FastDigest” que necesitan 10 minutos. La inactivación de la enzima se

realizó por calor en condiciones dependientes de la enzima, la mayoría durante 10 minutos a 65 °C.

2.5.6.Fosforilación y desfosforilación

La proporción de moléculas de vector recircularizadas en una ligación tras un proceso de digestión disminuye en gran medida si eliminamos los grupos fosfato terminales. Por ello, en esta tesis se defosforilaron los vectores utilizando las fosfatasas alcalinas de Fermentas “Shrimp alkaline phosphatase” y “FastAP thermosensitive alkaline phosphatase”. Se siguieron los protocolos del fabricante, en general una reacción de 20 µl de volumen con 1 pmol de DNA, 2 µl de tampón, 1 µl de enzima y agua “Milli-Q”, incubada 30min a 37 °C (10 min en el caso de “FastAP”).

Por otro lado, la ligación del vector desfosforilado con el inserto pertinente tendrá éxito siempre y cuando este tenga sus extremos fosforilados. En insertos procedentes de una digestión no es necesario, pero con moléculas procedentes directamente de una PCR se procedió a la fosforilación de los mismos. Para ello se utilizó “T4 polynucleotide kinase” de Fermentas, siguiendo su pertinente protocolo: 1-50 pmol de DNA, 2 µl de tampón, 2 µl de ATP 10 mM, 1 µl de enzima y agua Milli-Q hasta un volumen total de 20 µl incubado durante 20 min a 37 °C, e inactivando a 75 °C durante 10 min.

2.5.7.Reacciones de ligación

Las reacciones de ligación se realizaron con la “T4 DNA ligase” de Fermentas. El protocolo del fabricante indica preparar la siguiente mezcla: 20-100 ng de vector, inserto en una relación molar de entre 1:1 y 5:1 frente al vector, 1 µl de ligasa, 2 µl del tampón y agua hasta un volumen total de 20 µl. Si la ligación era entre extremos romos se añadió, además, 2 µl de polietilenglicol 4000 al 50%. Si bien se indica realizar la incubación 10 minutos a 22 °C, generalmente se incubó durante la noche. Se inactivó la enzima a 65 °C durante 10 min.

2.5.8.Transformación por electroporación

La transformación de las cepas con DNA se realizó mediante electroporación, dadas la alta eficiencia del método (Dower *et al.*, 1988) y las facilidades logísticas del laboratorio.

Para la preparación de las bacterias electrocompetentes se crecieron cultivos en LB hasta una OD₆₀₀ de 0,4-0,6, y tras exponerlas 30 min a hielo, se recogieron por centrifugación a 3000 g durante 15 min a 4 °C en una centrifuga “5810 R” de Eppendorf. A continuación, se realizaron 2 lavados en agua “Milli-Q” autoclavada y a 6 °C. Seguidamente se realizó un último lavado,

esta vez con glicerol 10% (v/v) también a 6 °C. Las centrifugaciones se realizaron a 3000 g durante 15 min a 4 °C. Después del último lavado, las células se resuspendieron en un volumen de glicerol 10 % de tal manera que se concentrasen 100 veces respecto al volumen original del cultivo. A continuación se realizaron alícuotas de 50 µl que fueron congeladas inmediatamente en un baño de hielo seco. Estas alícuotas se conservaron a -80 °C hasta su empleo.

Por su parte, el DNA tiene que estar libre de sales. El DNA procedente de una purificación se empleó directamente, pero el procedente de reacciones enzimáticas se microdializó antes de su uso. La muestra de DNA se dejó reposar durante 30 minutos sobre un filtro de nitrocelulosa “Millipore GS” de 0,05 µm de tamaño de poro, que a su vez flota en una placa de Petri con agua “Milli-Q”.

La transformación propiamente dicha se realizó mezclando una alícuota de células competentes y el DNA deseado. Cuando se trató de DNA plasmídico purificado se usaron 1-2 µl, mientras que en el caso de una ligación se usó todo el volumen de la reacción. Esta mezcla se añadió en una cubeta de electroporación de 0,2 cm “Gene Pulser” (BioRad), previamente enfriada en hielo. Se sometió a un pulso eléctrico usando un electroporador “MicroPulser” de BioRad utilizando el protocolo “Ec2” (2.5kV). Rápidamente esta mezcla fue resuspendida en 1 ml de medio LB estéril precalentado a 37 °C, se pasó el resultante a un tubo Eppendorf de 1,5 ml y se incubó a 37 °C en agitación durante 1-2 horas, en función de los genes de resistencia a expresar. Finalmente, se plaquearon 200 µl del cultivo en una placa de Petri con LB agar y los antibióticos adecuados para aislar los transconjugantes.

2.5.9.Método de ensamblaje isotérmico (*Isothermal assembly*)

Aparte de la clásica metodología de clonaje de corte con enzimas de restricción y ligación en esta tesis se han implementado procedimientos para la construcción de clones de reciente desarrollo. El ensamblaje isotérmico en reacción única de múltiples moléculas de DNA con extremos coincidentes (Gibson, 2011) resulta una metodología muy útil a la hora de construir dispositivos genéticos, y en el presente trabajo se utilizó para la construcción de plásmidos y la fusión de proteínas.

El método se basa en la siguiente sucesión de reacciones (Figura 2-1): La exonucleasa T5 elimina nucleótidos del extremo 5' de dos hebras dobles de DNA con secuencias terminales solapantes. De esta manera, estas dos moléculas presentan extremos terminales complementarios que se unen. Entonces la DNA polimerasa “Phusion” completa los huecos y

la ligasa *Taq* une los cortes. Toda la reacción se realiza a 50 °C; las 3 enzimas funcionan correctamente pero la exonucleasa se va inactivando a esa temperatura lo que evita que degrade más DNA del necesario.

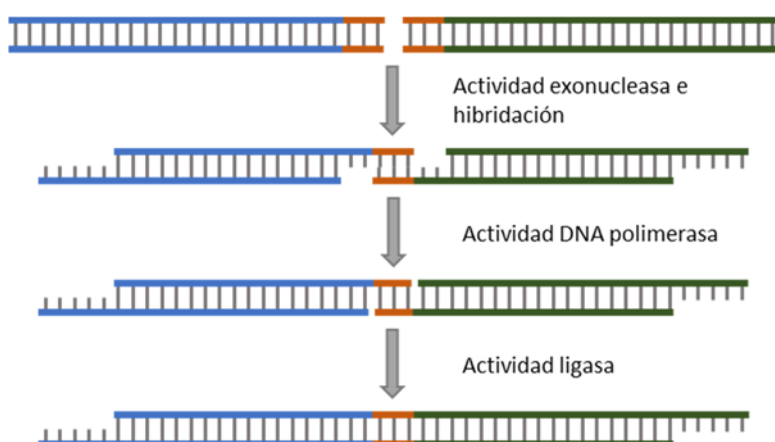


Figura 2-1. Proceso de ensamblaje de dos fragmentos de DNA. La exonucleasa elimina nucleótidos en la dirección 5'→3', dejando porciones de cadena sencilla, lo que hace que hibriden los extremos homólogos entre ambos fragmentos (en naranja). La DNA polimerasa adiciona nucleótidos en los extremos 3' libres y la ligasa sella las mellas.

Mediante este procedimiento se pueden ensamblar varios fragmentos de DNA, llegando incluso a cerrar una molécula circular, y por lo tanto, se puede construir un vector.

El procedimiento experimental llevado a cabo partió de PCRs para conseguir los fragmentos con los extremos cohesivos. Los amplificados se corrieron en gel y se purificaron las bandas adecuadas.

Se prepararon 6 ml de tampón mezclando 3 ml de 1M Tris-HCl pH 7,5, 150 µl de 2M MgCl₂ 60 µl 100 mM dGTP, 60 µl 100 mM dCTP, 60 µl 100 mM dTTP, 60 µl 100 mM dATP, 300 µl 1M DTT, 1,5 g PEG 8000, 300 µl 100 mM NAD y agua "Milli-Q" hasta el volumen total. Se dividió en 18 alícuotas de 320 µl, siendo congeladas a -20 °C para su posterior uso. Utilizando una de estas alícuotas se preparó la mezcla de reacción añadiendo 1,2 µl de exonucleasa T5 (Epicentre), 20 µl de polimerasa "Phusion" (Thermo), 160 µl de ligasa *Taq* (BioLabs) y 700 µl de agua "Milli-Q". Se prepararon alícuotas de 15 µl en tubos de PCR y se guardaron a -20 °C (siendo recomendable usarlas antes de un mes).

El siguiente paso fue mezclar en un volumen de 5 µl los distintos fragmentos a la misma concentración molar. Si se estaba construyendo un vector se utilizó como referencia el uso de aproximadamente 100ng del fragmento portador del origen de replicación y/o la resistencia antibiótica. Estos 5 µl se añadieron a una de las alícuotas de mezcla de reacción congeladas y se incubó la mezcla durante una hora a 50 °C.

Por último, el vector resultante se introdujo en la cepa deseada mediante electroporación, microdializándolo o purificándolo previamente.

2.5.10. Recombinación cromosómica con productos de PCR: método Datsenko-Wanner

Durante esta tesis fue necesaria la inserción de ciertas construcciones genéticas en el cromosoma bacteriano. Para ello se siguió un protocolo basado en el trabajo descrito por Datsenko y Wanner (Datsenko and Wanner, 2000). El método recombina un fragmento de DNA producto de una PCR en el genoma bacteriano. El lugar de recombinación se determina mediante unas secuencias en los extremos del amplicón que son homólogas a las secuencias donde se pretende que ocurra la recombinación. Para completar este proceso es necesario tener una cepa que exprese el sistema de recombinación λ Red, pudiendo estar codificado en un plásmido o en el propio cromosoma celular. En este trabajo se optó por la cepa TB10 que presenta el sistema de recombinación insertado en el genoma ya que tiene una eficiencia de 5 a 10 veces mayor que la opción del plásmido, aunque después de la recombinación es necesario realizar una transducción de la inserción a la cepa definitiva. La expresión de este sistema es dependiente de temperatura, se reprime a 30 °C y se induce a 42-45 °C.

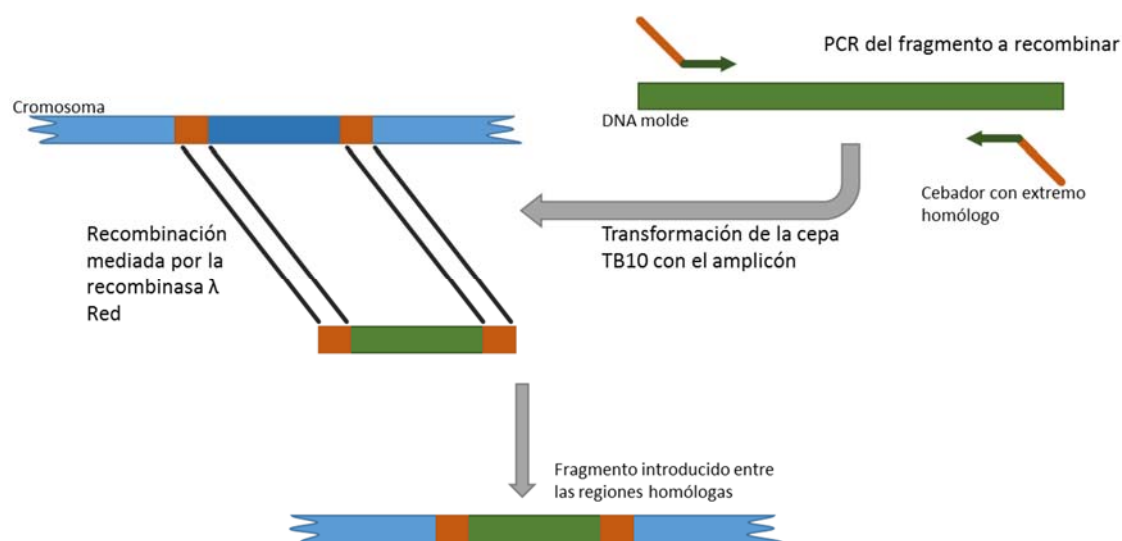


Figura 2-2. Proceso de recombinación empleado en la tesis. Mediante PCR se genera el fragmento de DNA que se quiere insertar en una determinada localización (representado por un rectángulo verde). En los extremos 5' de los cebadores se añade una secuencia de unos 50 nucleótidos (representado por líneas naranjas), que se corresponde con la región donde tendrá lugar la inserción. Cuando la cepa TB10, que contiene la recombinasa λ Red, es transformada con el amplicón, la recombinasa lleva a cabo una recombinación sitio específica que conduce a la inserción del amplicón en la localización deseada.

El primer paso de este método (Figura 2-2) fue la producción del fragmento a insertar en el cromosoma mediante una PCR. Dado que este DNA tiene que tener los extremos homólogos al cromosoma, es de principal importancia el diseño de unos cebadores con unas colas de aproximadamente 50 nucleótidos de homología con el lugar de inserción. Además, debe acarrear algún tipo de marcador antibiótico para aislar las células recombinantes. A

continuación, se digirió el amplificado con *DpnI* para eliminar posibles moléculas del molde y se purificó mediante electroforesis en gel para evitar fragmentos inespecíficos.

Se prepararon células electrocompetentes de TB10, siguiendo un protocolo similar al descrito previamente. La principal diferencia fue que el cultivo se realizó a 30 °C para silenciar la expresión del sistema de recombinación. Se partió de un cultivo saturado, diluyéndolo 1:70 y se creció hasta una OD₆₀₀ de 0,4. En ese momento se indujo la expresión del sistema de recombinación cambiando el cultivo a una estufa precalentada a 42 °C, incubándolo durante exactamente 15 min. El último paso fue lavar las células en agua fría estéril. En este caso, la conservación de las células para un posterior uso no es recomendable.

100 µl de células competentes se mezclaron con 10-100 ng de amplificado y se electroporaron, para a continuación resuspenderlas en 1 ml de medio LB o SOB. Se realizaron dos controles negativos en la electroporación: uno con células inducidas sin DNA, y otro con células sin inducir y DNA. Se incubaron durante 2-3 horas a 30 °C con agitación y se plaquearon 200 µl de los cultivos. El cultivo restante se plaqueó al día siguiente tras crecerlo durante la noche. Los recombinantes son especialmente sensibles a los antibióticos, por lo que las placas se prepararon a la mitad de concentración antibiótica habitual.

De las colonias obtenidas se realizaron dos siembras por estría consecutivos para garantizar el aislamiento de las cepas con la recombinación. Esta vez se utilizó la concentración antibiótica habitual. Finalmente se comprobaron las inserciones por PCR y secuenciación.

El último proceso de este método fue la transducción de la recombinación desde la cepa TB10 a la cepa BW27783 mediante el fago P1. Partiendo de un cultivo saturado de las células de TB10 recombinantes, se realizó una dilución 1:100 en 5 ml de LB suplementado con CaCl₂ 5mM y glucosa al 0,2 %. Se cultivó durante una hora a 30 °C y se inoculó con 100 µl de lisado de P1. Se incubó unas 3 horas hasta que el cultivo presentase una apariencia de lisado. Se añadieron 100 µl de cloroformo y se agitó con vortex. Esta mezcla se centrifugó y se guardó el sobrenadante, donde permanece el nuevo fago P1 con el DNA recombinante, a 4 °C.

El siguiente paso fue una segunda infección, esta vez con el P1 con el DNA deseado en la cepa BW27783. Se centrifugó 1 ml de cultivo saturado de BW27783 se resuspendió en 1 ml de LB más 100 mM MgSO₄ y 5 mM CaCl₂. Se prepararon 4 incubaciones: 100 µl de lisado P1 + 100 µl de células receptoras; 100 µl de lisado P1 diluido 1:10 + 100 µl de células receptoras; 100 µl de LB + 100 µl de células receptoras; y 100 µl de LB + 100 µl de lisado P1. Tras media hora de incubación a 37 °C se añadieron 200 µl de citrato de sodio (1 M, pH 5,5) y 1 ml de LB y se

incubó una hora más. Se centrifugaron las incubaciones, resuspendiéndolas en 180 µl de LB y 20 µl de citrato de sodio, y se sembraron en placas de LB agar con los antibióticos adecuados. Como en la infección anterior, de las colonias obtenidas se realizaron dos aislados consecutivos, esta vez añadiendo además citrato de sodio a las placas, y se comprobaron las inserciones por PCR y secuenciación.

2.5.11. Mutación sitio específica

En esta tesis fue necesario realizar una mutación puntual en la proteína GFP, para ello se adaptó la metodología “QuickChange” de Stratagene. Este método permite cambiar varios de nucleótidos de la secuencia de un plásmido, para ello sólo es necesaria una PCR de la que se obtiene directamente una molécula circular de DNA. Para ello es imprescindible el uso de una DNA polimerasa que no provoque desplazamiento de cadena, pero a pesar de lo indicado por los fabricantes es una propiedad que se manifiesta habitualmente. Por ello, en este trabajo se modificó dicho método, generando una molécula lineal de DNA con extremos romos que posteriormente fue ligada y convertida en circular (Figura 2-3).

El quid del método radica en el diseño de los cebadores para realizar una PCR con el cambio que se quiere realizar en su secuencia. En el presente trabajo solo uno de los oligonucleótidos llevaba la secuencia mutante, mientras que el cebador reverso era un oligonucleótido normal que hibridaba en el plásmido justo donde acababa el cebador directo. La molécula lineal resultante se sometió a una reacción de ligación y se trató con *DpnI* para eliminar el DNA molde. Finalmente, la molécula resultante se introdujo mediante electroporación en la cepa deseada.

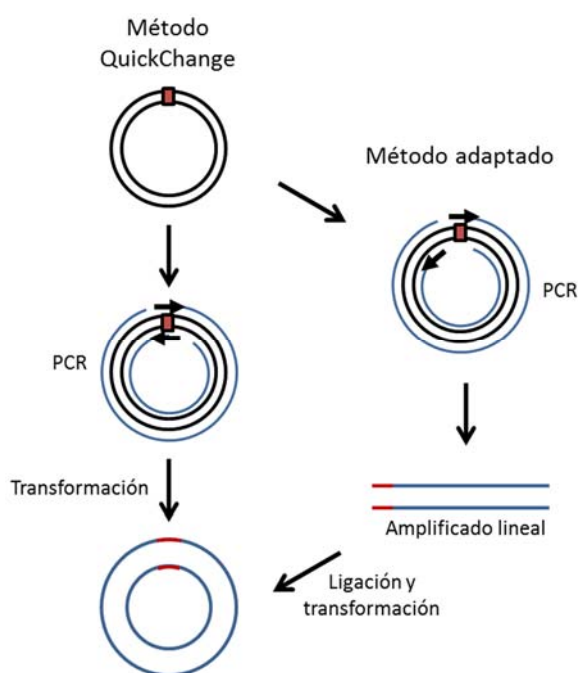


Figura 2-3. Mutagénesis dirigida mediante el método QuickChange. En el método tradicional QuickChange (izquierda) la mutación deseada (rectángulo rojo) se incorpora en dos cebadores complementarios. Estos se usan en una reacción de amplificación que genera una molécula circular con dos mellas, una en cada cadena y en posiciones diferentes. Las mellas son reparadas por la maquinaria celular. En la modificación del método (derecha de la figura), solo un cebador contiene la mutación y se genera un producto lineal que se liga *in vitro* y se transforma.

2.5.12. Secuenciación de DNA

La correcta construcción de todos los vectores y cepas de esta tesis fue verificada mediante secuenciación. Se utilizó el servicio de secuenciación Sanger de la empresa Stab Vida llamado "You Tube It". Los plásmidos fueron enviados directamente siguiendo las instrucciones de dicho servicio, mientras que para la comprobación de las cepas se realizaron PCRs de las regiones modificadas y dichos amplificados fueron enviados a secuenciar.

2.6. Construcción de dispositivos para el análisis de la regulación transcripcional

El planteamiento básico de los experimentos en esta tesis fue la medida y cuantificación del cambio de la actividad transcripcional de un promotor en función de la concentración de su proteína reguladora. Por lo tanto, su aplicación necesitó cuantificar la actividad promotora y a su vez controlar la cantidad de regulador de un cultivo en crecimiento.

La medición de la actividad transcripcional se realizó mediante fluorimetría de la GFP expresada por el promotor en un cultivo (Zaslaver *et al.*, 2006). Por otro lado, la cantidad de regulador transcripcional presente en el cultivo se controló usando el sistema de expresión del operón de la arabinosa. Este operón responde a dos señales para su activación, la arabinosa y el AMP cíclico (Schleif, 2000). La proteína reguladora AraC se comporta como un represor en ausencia de arabinosa, pero al unirse a esta molécula cambia su conformación e induce la actividad desde p_{BAD} . El cAMP, por su parte, actúa como coactivador al unirse a la proteína

reguladora CRP. En este estado CRP se une al promotor y aumenta su actividad. Además, se utilizó la cepa BW27783, un mutante para los transportadores de arabinosa en el que la concentración extracelular de la arabinosa es proporcional a la intracelular. Aprovechando estas características se pudo modular la cantidad de regulador producido añadiendo diferentes cantidades de arabinosa, se consiguió que concentraciones crecientes del inductor arabinosa en el medio supusieran crecientes niveles de expresión de regulador en la célula. Conjuntamente, se utilizó glicerol o glucosa como fuente de carbono para modificar los niveles de cAMP celular.

Lógicamente, el primer paso para el análisis fue la construcción de los sistemas que luego se iban a medir. Previamente al inicio de cualquier construcción, se planificó el proceso de clonaje y se diseñó la molécula de DNA en el programa “Vector NTI” de Invitrogen. En primer lugar se planearon dos plásmidos, uno sería el reportero y el otro el vector de expresión del regulador. Por un lado se amplificó el gen *tetR* del plásmido R100 con TetRKpnI y TetRHindIII como cebadores. En el oligonucleótido directo, TetRKpnI, se añadió además la secuencia 5’UTR presente en el vector pET-3a. Se digirieron el amplificado y el vector pBAD33 con las enzimas *KpnI* y *HindIII*, y ligando este inserto y el vector se consiguió el pRRG1. Asimismo se construyó pRRG58; se clonó el gen *gfp* en pBAD33, esta vez usando GFPKpnI y GFP2HindIII como cebadores. Por otro lado, se realizó otra PCR con los cebadores pTetA1XhoI y pTetA2BamHI y R100 como molde, amplificando la región promotora del gen *tetA*. Mediante las enzimas *XhoI* y *BamHI* se digirieron tanto el amplicón como el vector pUA66, y se ligó el inserto con el vector generando el vector pRRG13.

A continuación, se consideró necesario construir una fusión traduccional de TetR y mKate2. Se diseñó uniendo la proteína mKate2 al carboxilo terminal de TetR mediante un conector flexible compuesto por la cadena peptídica Ser-Gly-Gly-Gly-Gly (Chen *et al.*, 2012). Dada la complejidad de la estructura final se planeó su construcción mediante el ensamblaje isotérmico de tres fragmentos. Dos fragmentos fueron de producto de PCRs: una con los cebadores isomkatefusdir e isomkate_rev y pRAF22 como molde, amplificando el gen *mkate2*; y la segunda con los cebadores isotetRfusrev e isotetRfus_dir y R100 como molde, amplificando *tetR*. El tercer fragmento en cambio consistió en el vector pBAD33 linealizado mediante digestión por *XbaI*. Su ensamblaje dio como resultado el pRRG54. Siguiendo el mismo proceso, pero con los cebadores gfpfusion_dir y gfpfusion_rev para amplificar *gfp* e isotetRfus_dir y tetRgfpfusion_rev para amplificar *tetR*, se construyó el vector pRRG55.

Como en este último vector GFP estaba fusionada a TetR, hubo que construir un vector reportero con mKate2. Se realizó una PCR con los oligonucleótidos pCSmkate_dir y pCSmkate_rev con pRRG13 como molde. Por otro lado, se amplificó *mkate2* de pRAF22 con ptetAmkate_dir y ptetAmkate_rev. Su ensamblaje isotérmico resultó en el vector pRRG56.

Posteriormente se decidió bajar el número de copias del vector de expresión. Se diseñaron los oligonucleótidos pBADSEVA1 y pBADSEVA2 para amplificar la región de expresión de pBAD33, desde *araC* hasta el terminador. De esta manera, se pudo amplificar cualquier clon realizado previamente en el MCS de pBAD33. Tras digestión por *PacI* y *SpeI* y ligación del inserto y del vector pSEVA121, se consiguió el vector con menor número de copias. De esta manera se construyeron, entre otros, pRRG9, pRRG57, pRRG62 y pRRG68.

En un siguiente paso se decidió unificar el sistema reportero y el de expresión en un mismo vector (Figura 2-4). Se amplificaron las regiones de los vectores reporteros que contenían el promotor, la proteína fluorescente y la resistencia antibiótica y se clonaron en los vectores de expresión de bajo número de copias correspondientes. Además mediante esa PCR se añadió el terminador Bba_B1007, que quedaría en mitad de las dos regiones aislándolas. Como ejemplo, en el caso de la construcción de pRRG63, se amplificó la región reportera de pRRG13 con pUA66terB1007 y ptetAGFPKm_revPacI y se introdujo en el pRRG62 mediante corte con *PacI* y ligación.

Mención aparte merece pRRG74, plásmido diseñado como indicador de la actividad del par regulador/promotor en el operón *tet*. Se diseñó conservando la configuración y secuencia que presentan *tetR*, p_{tetR} y p_{tetA} en Tn10. Para cuantificar TetR se fusionó a mKate2 y GFP se utilizó de reportera de la actividad de p_{tetA}. Siguiendo el diseño se amplificó: el gen *mkate2* con los cebadores pRRG74_mkatedir y pRRG74_mkaterrev; el gen *gfp* con pRRG74_GFPdir y pRRG74_GFPprev; y la región del operón *tet* de Tn10 con pRRG74_ptetArev y pRRG74_TetRev. El ensamblaje de estos tres fragmentos generó el vector pRRG74.

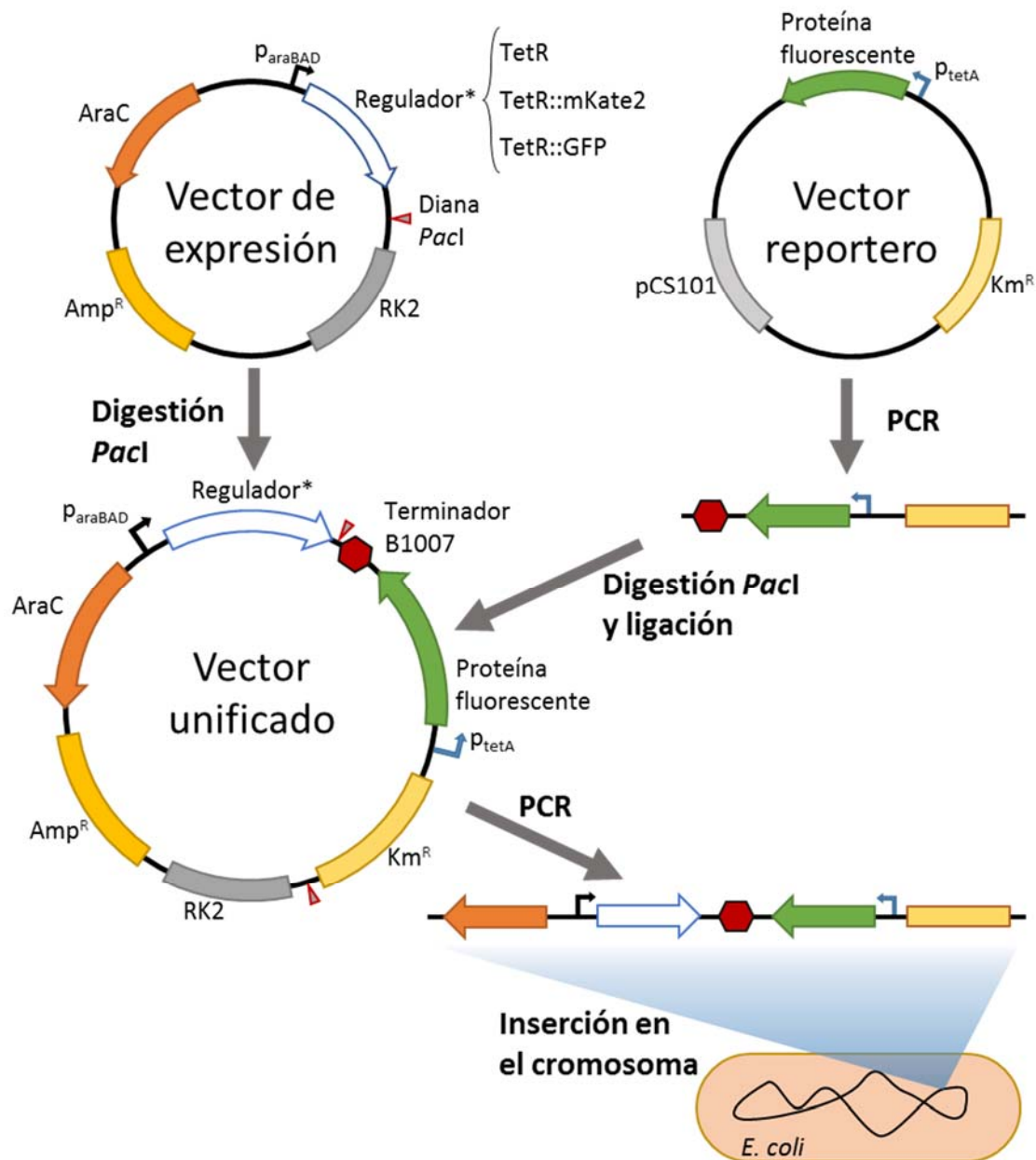


Figura 2-4. Representación de los dispositivos genéticos empleados y proceso de construcción del sistema unificado en plásmido y en cromosoma. Los orígenes de replicación de cada vector se representan con cajas grises; los genes de resistencia a antibióticos, con cajas amarillas; los genes de reporteros fluorescentes, en verde; el represor de estudio, por una flecha blanca con borde azul. La región que contiene el gen reportero bajo el promotor p_{tetA} y el marcador de resistencia Km^R se amplificó por PCR a partir del vector reportero, con extremos digeribles por el enzima de restricción $PacI$. Dicho fragmento se insertó en el único sitio $PacI$ del vector de expresión, dando lugar al vector unificado. Mediante PCR se obtuvo un fragmento que contenía, tanto la región reportera previamente integrada, como el sistema $araC - p_{BAD} -$ gen represor de p_{tetA} . Este *cassette* represor-reportero se insertó de manera sitio-específica en el cromosoma.

Todos estos vectores se metieron en *E. coli* BW27783 mediante electroporación, generando las cepas con las que se analizaría la actividad transcripcional.

Como último planteamiento, se decidió introducir las construcciones en el cromosoma de *E. coli* BW27783, en el locus del operón de la arabinosa. Para ello, se realizaron PCRs que amplificaban las regiones reporteras y de expresión de los plásmidos unificados, y siguiendo el

protocolo descrito en la sección 2.5.10, se recombinaron en TB10 y posteriormente se transdujeron a BW27783.

2.7. Cuantificación de la actividad transcripcional mediante fluorímetro

La actividad de un promotor en función de la cantidad de su regulador transcripcional en un cultivo en crecimiento se midió siguiendo el método detallado en (Fernandez-Lopez *et al.*, 2010).

Los cultivos que contenían el sistema de cuantificación se crecieron en M9 suplementado con casaminoácidos (0,2%) y glicerol o glucosa (0,5%) en presencia de los antibióticos adecuados hasta fase de saturación. A continuación, se diluyó el cultivo 1:10000 y 150 μ l se añadieron a los pocillos de una placa *microtiter* de 96 pocillos transparente (Deltalab)(Figura 2-5.A). En función del experimento a realizar, se generaron gradientes de arabinosa a lo largo de las filas, habitualmente de 16 a 0 μ M. Dado el número de pocillos, diferentes cultivos y controles pudieron ser analizados en una misma placa.

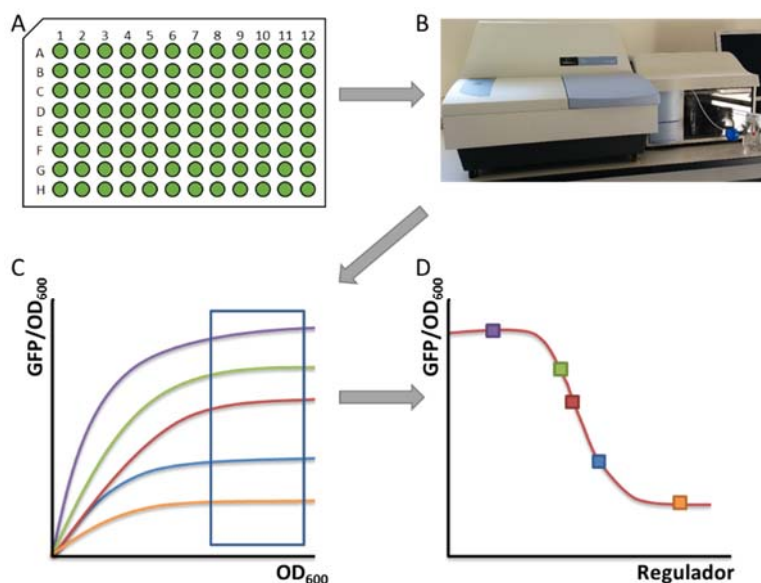


Figura 2-5. Medición de la actividad transcripcional en fluorímetro. (A) Placas de 96 pocillos se utilizan para realizar medidas de distintos cultivos y condiciones al mismo tiempo. (B) La medición y el cultivo se realiza en un fluorímetro Victor³. (C) Se obtienen valores de fluorescencia y absorbancia. (D) El promedio del estado estacionario representa la actividad promotora para cada condición.

Estas placas se incubaron durante 14 horas en un fluorímetro Victor3 (PerkinElmer) a una temperatura de 37 °C con agitación orbital (Figura 2-5.B). La fluorescencia y la absorbancia fueron medias cada 7 minutos para cada pocillo. Dada la temperatura de incubación, el medio se evaporaba, de modo que hubo que contrarrestarla añadiendo agua cada tres ciclos de medida.

Los datos de absorbancia no se correspondían con la habitual escala de OD₆₀₀ por lo que se les restó el fondo obtenido de lecturas de medio sin células y se transformaron mediante una curva de calibración obtenida mediante valores del fluorímetro y un espectrofotómetro (Shimadzu UV-1603). Las medidas de fluorescencia también presentaban una lectura de fondo. En este caso la propia cepa presentaba autofluorescencia, por lo que a los valores medidos se les restó el valor equivalente a un cultivo sin GFP, en su mismo estado de crecimiento en función de la OD₆₀₀. Se relativizó la fluorescencia frente al crecimiento (GFP/OD₆₀₀)(Figura 2-5.C) y se contabilizaron los valores en la fase exponencial de crecimiento celular, siendo su media la actividad promotora en cada cultivo (Figura 2-5.D).

2.8. Cuantificación de la actividad transcripcional mediante microscopía de fluorescencia

Las construcciones para el estudio de la actividad transcripcional también se analizaron mediante microscopía de fluorescencia. En este caso, en lugar de medir las células a lo largo de su crecimiento se optó por hacer medidas puntuales de cultivos en fase exponencial para cada condición de interés, es decir en presencia de glicerol o glucosa y a distintas concentraciones de arabinosa.

2.8.1. Adquisición de imágenes

Cultivos saturados en M9 con los antibióticos y las condiciones deseadas se diluyeron 1:1000 en el mismo medio y se dejaron crecer 2-3 horas. En ese momento, se recogieron 2 µl y se depositaron sobre un *pad* de arabinosa. Estos *pad* se prepararon pegando un marco “Frame-Seal Incubation chamber” (BioRad) sobre un portaobjetos (VWR), formando una cavidad de 65 µl. En ella se añadieron 200 µl de M9 suplementado con arabinosa al 1,5 % (con las mismas condiciones en las que el cultivo correspondiente hubiese crecido). Fue necesario realizar este paso con el medio caliente, para que la arabinosa estuviese en estado líquido. Inmediatamente después de añadir M9 agarosa en el centro del espacio dejado por el marco, se depositó otro portaobjetos sobre el primero, para conseguir un gel nivelado y liso. Tras media hora dejándolo reposar y solidificar se retiró con cuidado el portaobjetos superior. Usando un bisturí se cortó un cuadrado de unos 0,25 cm² del gel de agarosa. Este cuadrado se mantuvo en el portaobjetos, retirándose la agarosa restante. En este cuadrado se depositaron los 2 µl de cultivo, y se dejaron secar unos 5 minutos, tras los cuales se colocó un cubreobjetos (24x24 mm N°1 de VWR) sobre el marco, sellándolo con precaución.

Las muestras, preparadas en los *pads* sellados, se observaron en un microscopio invertido de epifluorescencia Leica AF6500. El microscopio contaba con cabina de incubación, de tal manera que las muestras siempre permanecieron en crecimiento a 37 °C.

Las imágenes se adquirieron con un aumento de 63x mediante el objetivo “HCX PL S-APO 63x 1,30 Oil”. Se adquirieron imágenes en contraste de fases y en dos canales de fluorescencia, roja y verde. Se utilizaron los filtros correspondientes para la longitud de onda de emisión y excitación de las proteínas fluorescentes: (Ex. 562/40 - Em. 641/75) para mKate2, y (Ex. 482/18 - Em. 520/28) para GFP. Asimismo, se aplicaron diferentes tiempos de exposición en función de la intensidad de la fluorescencia de las muestras. La fuente de luz externa EL6000 de Leica, equipada con una lámpara de vapor de mercurio (HXP Short Arc Lamp de Osram), fue utilizada para la excitación de la fluorescencia. Se empleó a su máxima intensidad. Además, debido a que los fluoróforos se degradaban con la luz, se conservó una distancia de 4 campos entre una posición de toma de imagen y la siguiente. Las imágenes se obtuvieron mediante la cámara de alta velocidad de 12bits Andor iXon885.

Todo el manejo del equipo y la adquisición de imágenes se realizaron desde un ordenador con el programa “LAS AF” de Leica.

2.8.2.Tratamiento de imágenes

Las imágenes obtenidas se exportaron a formato “tif” y se analizaron usando el software MATLAB de MathWorks. A continuación se procedió a cuantificar la intensidad de fluorescencia en cada canal para cada célula. Para ello se llevaron a cabo varios pasos:

Corrección de campo plano

Debido a las propiedades de las lentes la luz no llega de manera homogénea a todos los píxeles de la cámara. Esto ocurre con una curvatura que presenta su máximo en el centro del campo. Se realizaron varias imágenes de campos vacíos con las configuraciones del microscopio y la cámara que se utilizaron con las muestras. Estas imágenes vacías sirvieron como descriptor de esta aberración y realizando una media de todas ellas se obtuvo una matriz de corrección. Aplicando esta matriz a las imágenes a analizar se consiguió una señal homogénea en todo el campo (Figura 2-6).

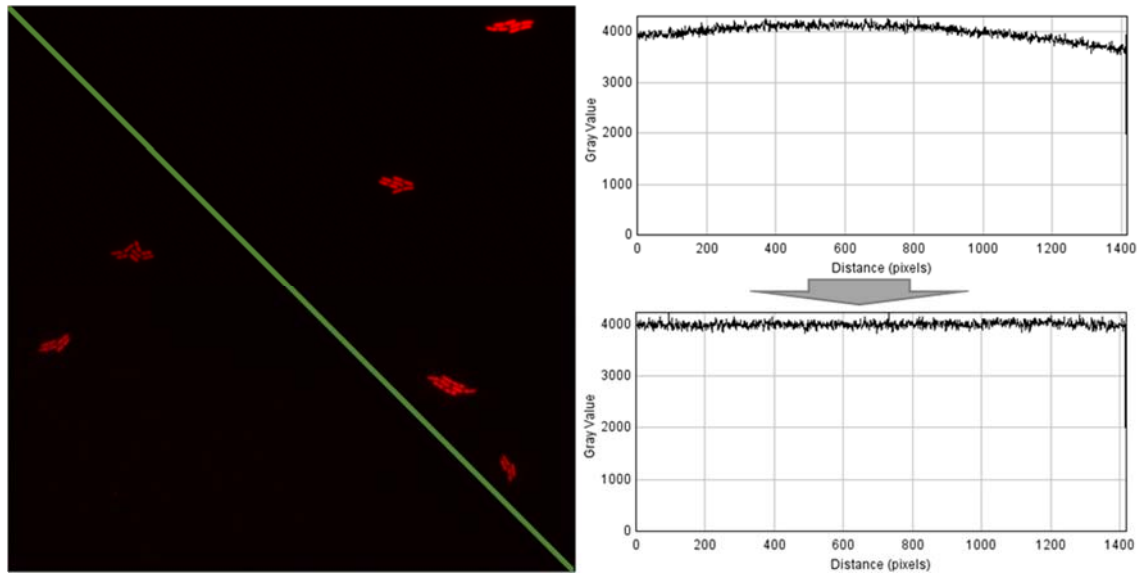


Figura 2-6. Corrección de campo plano. Se observa cómo el perfil de la imagen cambia de curvado a plano tras la corrección. A la izquierda se muestra una imagen típica de fluorescencia en el canal del rojo. Se trazó una línea indicada en verde en la figura, y se obtuvieron los valores de fluorescencia para cada punto ubicado sobre la línea. En la parte superior derecha, se representan la intensidad de los puntos ubicados sobre la línea verde en relación con la distancia a la que se encuentran de la esquina superior izquierda. En la parte inferior derecha, se muestran los valores obtenidos tras aplicar la matriz de corrección.

Segmentación

A definir qué área de una imagen corresponde a cierto objeto se le llama segmentación. En este trabajo se utilizó el programa “MicrobeTracker” (Sliusarenko *et al.*, 2011) para delimitar el área que correspondía a cada bacteria en las imágenes de contraste de fases. Se configuraron los parámetros del programa para obtener la mejor calidad de segmentación. Aun así, el programa cometía muchos errores con nuestras imágenes por lo que todas las segmentaciones automáticas hubo que corregirlas y depurarlas manualmente. Así se obtuvieron las llamadas mallas de segmentación, es decir, el conjunto de áreas que definen las células. Al mismo tiempo, utilizando dichas mallas en los demás canales, se computó el nivel de fluorescencia de cada célula, sumando las intensidades de los píxeles dentro del área delimitada (Figura 2-7). Además, se detectó y restó la señal de fondo con una herramienta que lleva el propio “MicrobeTracker”.

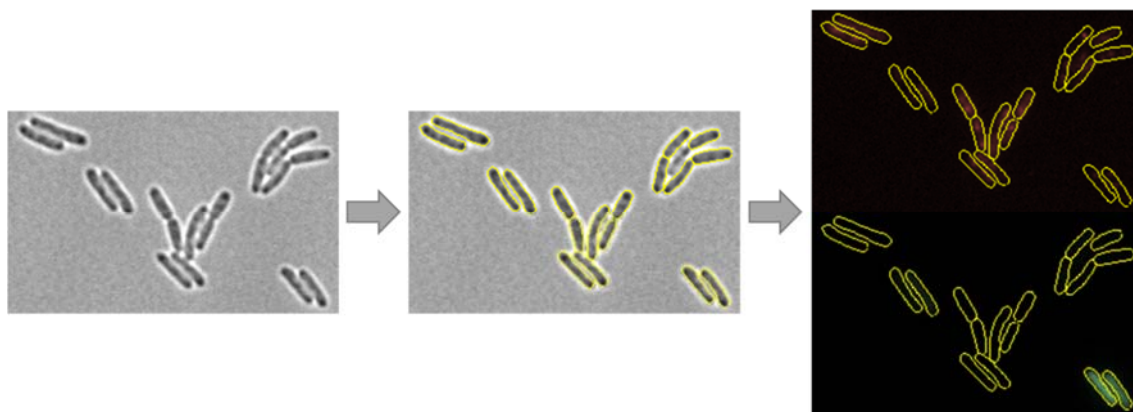


Figura 2-7. Segmentación con MicrobeTracker. La imagen ubicada en la izquierda de la figura muestra un campo obtenido en contraste de fases. La imagen central indica la segmentación automática obtenida tras aplicar MicrobeTracker, con las células contorneadas en amarillo. Las imágenes de la derecha muestran las máscaras de segmentación sobre el canal del rojo (superior) y del verde (inferior).

Desplazamiento de la fluorescencia

Las distintas longitudes de onda atraviesan de manera distinta la óptica de un microscopio, debido a la refracción, por lo que se producen efectos de desplazamiento entre los distintos canales analizados. Esto es un problema si vamos a tomar un canal como referencia a la hora de analizar dónde están los objetos en los demás canales. En este estudio se observó que la imagen de las bacterias en contraste de fase no se superponía perfectamente con su imagen en los canales de fluorescencia. Contrariamente a lo esperado, el desplazamiento no era sistemático, no en todas las células ocurría ni en la misma dirección ni en el mismo grado. Esto ocurría debido no solo a la refracción, sino también a imperfecciones en la superficie de los *pads* y a las diferentes disposiciones de las células. Al no seguir un patrón la solución resultaba excesivamente compleja. Se analizó la pérdida de señal de las células en los distintos canales, comparándola respecto a un análisis corregido manualmente (Figura 2-8). Se llegó a la conclusión de que la pérdida de señal era proporcional y constante en todos los canales, de tal manera que la relación entre las intensidades de las distintas fluorescencias no variaba. Por lo tanto, y dada la complejidad de la corrección, se optó por no realizarla.

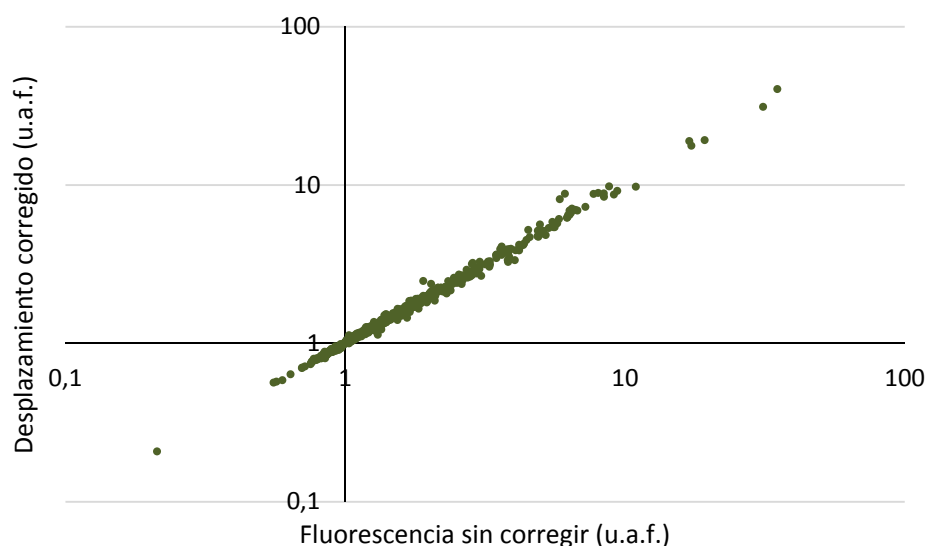


Figura 2-8. Correlación de la fluorescencia cuantificada con el desplazamiento corregido y sin corregir. En el eje horizontal se muestra la fluorescencia obtenida mediante la segmentación automática de MicrobeTracker sin corregir en el canal verde de fluorescencia. En el eje vertical se representa la fluorescencia obtenida tras la corrección manual del desplazamiento. Se muestran los datos para 750 células con el vector pRRG63.

Escalado

El rango de intensidades de fluorescencia que fue necesario medir no se pudo caracterizar con un mismo tiempo de exposición para todos los casos, ya que algunos valores quedaban por debajo del nivel de detección mientras que otros saturaban la señal. Consecuentemente, se aplicaron diferentes exposiciones para los diferentes niveles de actividad esperados. Para poder unificar en una misma escala los valores obtenidos con distintos tiempos de exposición se midieron como controles unas esferas con fluorescencia predefinida (“Rainbow Fluorescent Particle Slide” de Spherotech). Con los datos dados por estas esferas, se calculó la relación entre los valores de las distintas exposiciones (Figura 2-9). Además, se observó que relativizando el valor de fluorescencia frente al tiempo de exposición los valores se correspondían entre tiempos de exposición diferentes en menos de un orden de magnitud, aplicándose posteriormente para el análisis de datos de las bacterias en crecimiento.

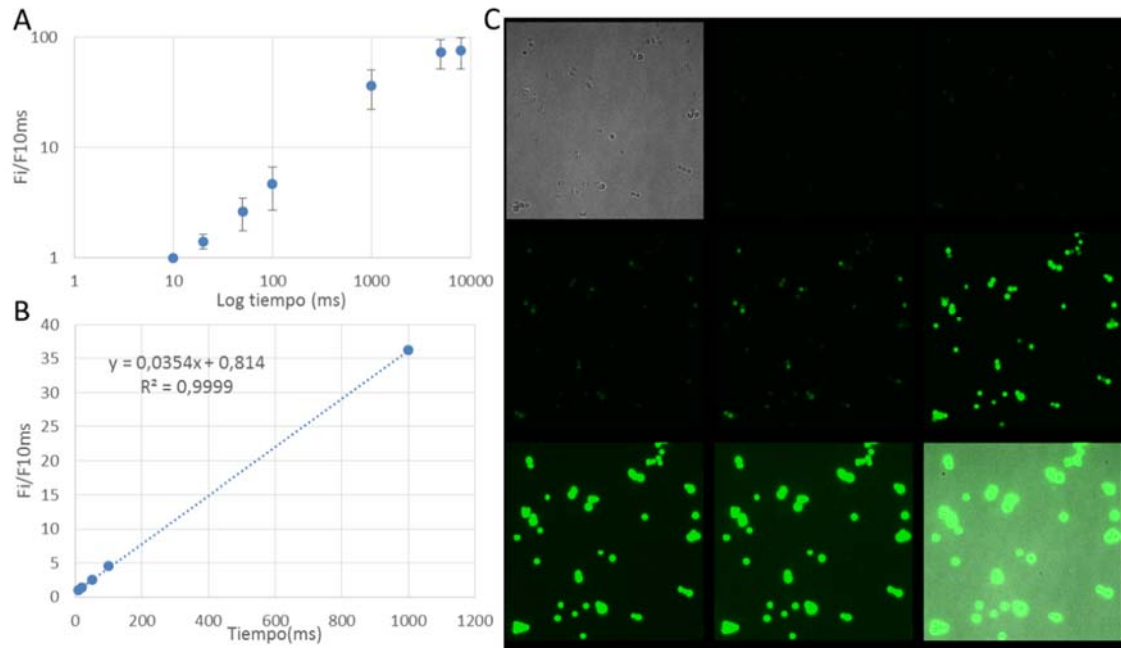


Figura 2-9. Control de los niveles de detección de fluorescencia y su escalado con las esferas de calibración. (A) El eje horizontal muestra el logaritmo en base 10 del tiempo de exposición en ms. El eje vertical muestra los valores de fluorescencia obtenidos para cada exposición normalizados por el valor obtenido a los 10 ms. (B) Regresión lineal de los valores representados en (A). (C) Nueve imágenes de un campo. La primera corresponde a la imagen de contraste de fases. Las siete siguientes corresponden a tiempos de exposición de 10, 20, 50, 100, 1000, 8000 y 10000 ms. La última imagen (inferior derecha) corresponde a la superposición de las ocho imágenes anteriores.

2.9. Simulaciones y análisis computacional

2.9.1. Ecuaciones maestras

Dentro de la célula, las moléculas que regulan los procesos de transcripción y traducción están presentes en bajo número de copias, lo que produce un alto grado de estocasticidad en la síntesis y degradación de mRNAs y proteínas (Elowitz *et al.*, 2002; Ozbudak *et al.*, 2002; Paulsson, 2004). Los modelos clásicos que utilizan ecuaciones diferenciales ordinarias (ODEs) asumen concentraciones continuas de los componentes, y no son, por tanto, adecuados para capturar la dinámica probabilística de la transcripción y la traducción. La forma más directa de abordar estos procesos consiste en considerar que mRNAs y proteínas varían sus concentraciones de forma discreta, en unidades o múltiplos de unidades (Ozbudak *et al.*, 2002; Paulsson, 2004). Para describir esta situación, consideramos al sistema de estudio como compuesto de n dimensiones, donde n es el número de especies analizadas. Si analizamos, por ejemplo, la dinámica de un mRNA y su proteína correspondiente, utilizaremos dos especies (x_1 para el mRNA y x_2 para la proteína, por ejemplo). Describimos el sistema asignando valores a cada una de las dimensiones, de tal forma que si existen c posibles valores para cada una de las especies, el sistema en cada momento del tiempo toma un valor de los c^n estados posibles. Para representar un sistema de mRNA y proteína utilizaríamos, por tanto, una matriz X_{ij}

donde el subíndice i nos informaría de la concentración de mRNA y el subíndice j de la concentración de proteínas (Figura 2-10.A). Puesto que el sistema tiene un número finito de estados, a cada uno de estos estados se le puede asignar una probabilidad $P(X_{ij})$, y lo que nos interesa es el cambio en esa probabilidad con respecto al tiempo $dP(X_{ij})/dt$. A las ecuaciones que describen la variación temporal de probabilidad de un sistema de encontrarse en un estado dado, las denominamos *ecuaciones maestras* (CMEs) (Van Kampen, 2007), y constituyen la principal herramienta teórica que utilizaremos para desarrollar los modelos computacionales en esta tesis.

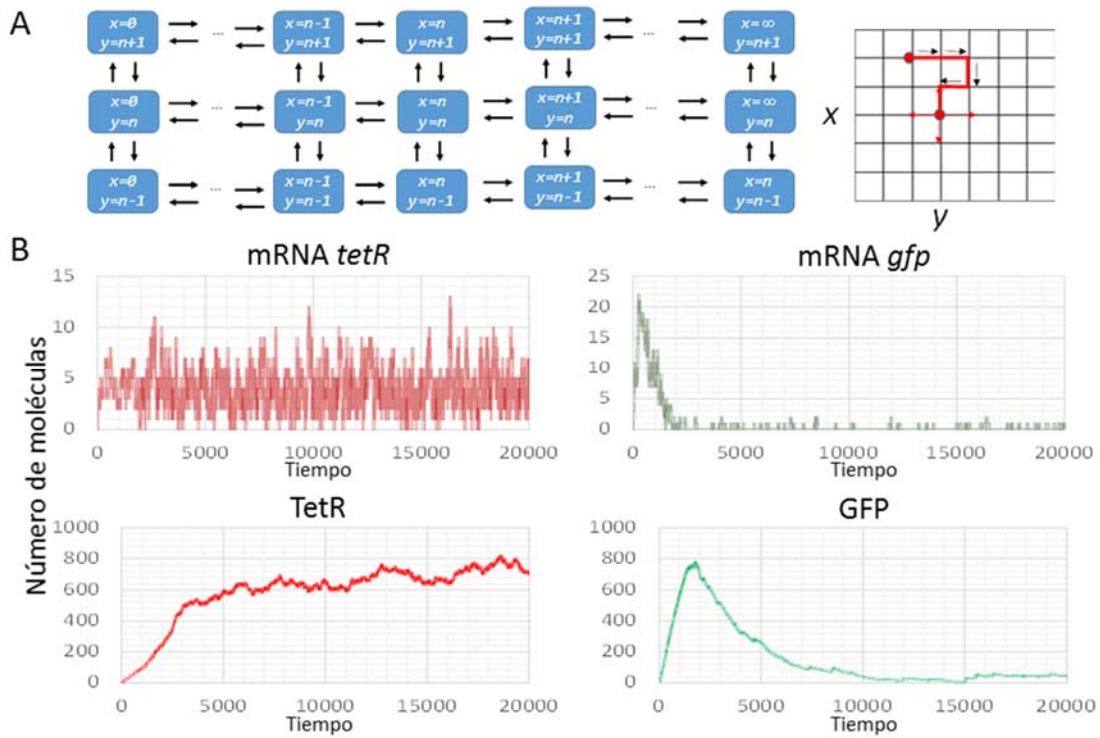
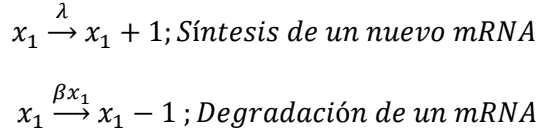


Figura 2-10. (A) Esquema de un proceso de Markov en dos dimensiones (x, y). Cada rectángulo representa un estado posible del sistema. Las flechas indican las transiciones elementales entre estados. Representado sobre el plano de fases (derecha), un proceso de Markov en dos dimensiones toma la forma de un paseo aleatorio sobre una malla. Durante un tiempo t el sistema progresa desde su estado inicial al estado final (puntos rojos) a través de transiciones discretas (flechas negras). Una vez en un estado dado, las posibles transiciones del sistema (flechas rojas), son independientes de la trayectoria que el sistema siguió hasta llegar a ese punto. (B) Trayectorias del sistema de CMEs (Ecuación 2-11) correspondientes a un regulador y su promotor regulado.

Una CME constituye una ecuación diferencial en el tiempo para la probabilidad del sistema de encontrarse en un estado dado. A diferencia de los modelos de ODEs, las CMEs consideran tiempo y probabilidad como las únicas variables continuas y diferenciables. Se expresan por medio de reacciones elementales: aquellos procesos fundamentales que producen un cambio en el sistema. Por ejemplo, supongamos un mRNA que se produce a una tasa constante (un promotor no regulado) y se degrada de forma convencional, siguiendo una cinética de primer orden. En ese caso escribiríamos:



Ecuación 2-1. Reacciones elementales de la CME para el cambio en la concentración de un mRNA.

Esta CME indica que los mRNAs se producen de uno en uno con una tasa constante, λ , y se degradan por medio de una cinética de primer orden con una tasa βx_1 . A partir de estas reacciones, podemos expresar la CME que describe la probabilidad de que el sistema se encuentre con una concentración determinada de mRNA:

$$\frac{dP(x)_t}{dt} = \lambda P(x-1)_{t-dt} + \beta x_1 P(x+1)_{t-dt} - \lambda P(x)_{t-dt} - \beta x_1 P(x)_{t-dt}$$

Ecuación 2-2. CME para la concentración de mRNA.

Esto es, el cambio en probabilidad de que la célula tenga una concentración de mRNA x_1 en el momento t corresponde a:

- La probabilidad $P(x)$ de que el sistema *entre* en esa concentración particular x en el intervalo de tiempo dt . Esto es igual a la tasa de síntesis λ , multiplicado por la probabilidad de que el sistema se encuentre en la concentración $x-1$ (puesto que la síntesis produce una molécula más y $x-1+1=x$), sumado a la tasa de degradación $\beta(x+1)$ multiplicado por la probabilidad de que el sistema se encuentre en el estado $x+1$. Esto es, para cada valor de x , el sistema puede llegar a él bien produciendo una molécula desde el estado $x-1$, bien degradando una molécula desde el estado $x+1$.
- A la probabilidad de que el sistema *entre* en x desde $x-1$ o desde $x+1$ hay que restarle la probabilidad de que el sistema ya se encontrase en el estado x , y salga de él, bien por síntesis ($\lambda P(x)$), bien por degradación ($\beta x P(x)$).

En general, cuando consideramos una única dimensión o variable de estado, una CME adquiere la forma:

$$\frac{dP(x)_t}{dt} = \sum_k r_k P(x-s_k) - \sum_k r_k P(x)$$

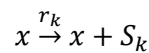
Ecuación 2-3. CME para una variable o estado.

Esto es, si tenemos k reacciones, cada una de ellas con una tasa r_k y produciendo un número de moléculas s_k , el cambio en la probabilidad de que el sistema se encuentre en el estado x es igual a la suma de todos los flujos de entrada (desde los estados apropiados $x-s_k$) menos la suma de todos los flujos de salida desde el propio estado x . Las tasas de cambio r_k pueden ser funciones explícitas del tiempo t , pero a menudo r_k es una función únicamente del estado del sistema. Decimos entonces que el sistema experimenta *transiciones sin memoria*, esto es, la

probabilidad del sistema de cambiar del estado x a un estado $x + s_k$ no depende de la trayectoria que el sistema experimentó hasta llegar a x . A este tipo de procesos se le denomina *Procesos de Markov* (Van Kampen, 2007).

2.9.2. Procesos de Markov y algoritmo de Gillespie

Los modelos estocásticos de los procesos de transcripción/traducción asumen que la producción y la degradación de las moléculas participantes (mRNAs y proteínas) pueden representarse como reacciones elementales (Paulsson, 2004):



Ecuación 2-4

En estas reacciones, los componentes del sistema cambian en forma de números enteros S_k de acuerdo a las tasas r_k . Tanto r como S_k son específicos para cada reacción k del sistema. Con frecuencia $S_k=1$, de tal forma que las variables de estado cambian de unidad en unidad, aunque S_k puede tomar valores mayores de 1 (*burst size*, o tamaño de la ráfaga) cuando una misma reacción produce o degrada varios componentes al mismo tiempo (por ejemplo, en la dimerización de dos proteínas la concentración de monómero decrece con $S_k=2$). En general, lo que justifica el tratamiento de transcripción y traducción como reacciones elementales no es tanto la bioquímica del proceso (la síntesis de un mRNA o una proteína implica un proceso de elongación e incorporación de nucleótidos/aminoácidos) como su cinética. La condición fundamental de un proceso de Markov es su falta de memoria, condición que cumplen aquellos procesos que muestran tiempos de espera distribuidos exponencialmente (Howard, 1971). Dado que tanto traducción como transcripción están típicamente controlados por un paso limitante (en bacterias, generalmente a nivel de la iniciación de ambos procesos), lo esperable es que la mayoría de los mRNAs y proteínas bacterianos experimenten tiempos de reacción exponencialmente distribuidos. De hecho, la mayoría de los estudios experimentales en bacterias han demostrado distribuciones poissonianas o binomiales negativas para mRNA y proteínas, lo que indicaría tiempos de espera exponenciales con $|S_k|=1$ (Poisson) o $|S_k|>1$ (Binomial negativa).

Para la resolución numérica de una CME disponemos de un algoritmo capaz de generar trayectorias de manera exacta. El algoritmo de Gillespie pertenece a los denominados métodos de Monte Carlo dinámicos (Gillespie, 1977). Aunque computacionalmente intensivo, Gillespie permite dar soluciones numéricas exactas a la CME. Para ello el algoritmo realiza dos operaciones básicas. La primera consiste en elegir los tiempos de reacción, esto es, determinar la distribución de tiempos de espera entre reacción y reacción. La segunda operación básica

consiste en, una vez que se ha determinado el tiempo de reacción, elegir de entre todas las reacciones posibles cuál sucederá de acuerdo a sus propensiones (que corresponden a las tasas instantáneas para cada reacción). De esta forma, si tenemos un sistema con n especies distintas, tendremos n variables de estado: $x_1, x_2, \dots, x_n$. Estas n variables de estado pueden experimentar K reacciones distintas, con tasas $r_1, r_2, \dots, r_K$. Una reacción i de las K posibles añade S_{ij} moléculas (donde S_{ij} puede ser negativa o positiva) a la variable de estado x_j . El algoritmo de Gillespie consiste en:

1.- Calcular la tasa total de reacción, esto es, la suma de las tasas de todas las reacciones posibles del sistema. $r_t = \sum_{i=1}^K r_i$, donde r_i puede ser una función arbitraria de las variables de estado $r_i = f(x_n)$. Por ejemplo, si un mRNA se sintetiza con tasa constante λ y se degrada en una cinética de primer orden βx_i , la tasa total cuando la célula contiene 10 mRNAs sería $r_t = \lambda + 10\beta$. Esta tasa total nos indica cuánto tiempo esperamos entre reacción y reacción. Para un proceso sin memoria y con tiempos de espera exponencialmente distribuidos, este tiempo correspondería a $F(t) = 1 - e^{-r_t t}$

2.- Una vez que se ha determinado el tiempo de reacción, se elige cuál de las reacciones tendrá lugar de acuerdo a su probabilidad $\frac{r_i(x)}{r_t(x)}$. Si la reacción i es elegida, entonces se actualizan las concentraciones de las variables de estado x_n de acuerdo a S_{ij} . Una vez hecho esto, el algoritmo itera al paso 1.

El algoritmo de Gillespie es una solución numérica exacta de la CME

Para la ejecución del algoritmo de Gillespie, lo primero es elegir un número aleatorio de distribución uniforme u_t con valores entre 1 y 0. Este número aleatorio se transforma entonces en un número exponencialmente distribuido del cual extraemos el tiempo de reacción:

$$u_t = 1 - e^{-r_t t}$$

Ecuación 2-5

De tal forma que:

$$t = -\frac{1}{r_t} \ln(1 - u_t)$$

Ecuación 2-6

Una vez elegido el tiempo de reacción, se determina cuál de las K reacciones sucede. Para ello, se genera un segundo número aleatorio u_t con valores entre 1 y 0. Normalizamos cada una de las tasas individuales r_i por r_t de manera que si $u_t \leq r_1/r_t$ se elige la reacción r_1 . Si por el

contrario $u_t > r_1/r_t$, se comprueba si $u_t \leq (r_1 + r_2)/r_t$, y en caso afirmativo se elige r_2 . El proceso se repite hasta que $u_t \leq (r_1 + r_2 \dots + r_i)/r_t$, eligiéndose r_i .

Gillespie presentó su algoritmo como un método de simulación para especies químicas en un medio bien mezclado, asumiendo difusión perfecta (Gillespie, 1977). Por esta razón, se suele aducir que el algoritmo no es adecuado para simular reacciones de componentes con difusión limitada (como macromoléculas) o que se encuentren en medios heterogéneos. Sin embargo, esto no es cierto (Paulsson, 2004). El algoritmo de Gillespie no es más que una reformulación de la CME en términos numéricos para cualquier proceso dinámico markoviano de tipo:

$$A \xrightarrow{r} B$$

Dado que este es un proceso dinámico, estamos interesados en determinar la probabilidad de que el sistema esté en A o en B en un instante de tiempo $t + dt$, de tal forma que:

$$P(A, t + dt) = P(A, t) - P_r(t, t + dt)$$

Ecuación 2-7

Esto es, la probabilidad de que el sistema permanezca en A hasta el momento $t + dt$ corresponde a la probabilidad de que ya estuviera en A en el momento t , menos la probabilidad de tránsito P_r en el tiempo dt . Utilizando la ley de la probabilidad total podemos condicionar esta probabilidad de tránsito:

$$P_r(t, t + dt) = P_r(t, t + dt|A, t)P(A, t) + P_r(t, t + dt|B, t)P(B, t)$$

Ecuación 2-8

Sin embargo, la segunda parte de esta expresión es cero, porque la probabilidad de tránsito en el caso de que el sistema ya se encuentre en B es cero. De tal forma que

$$P_r(t, t + dt) = P_r(t, t + dt|A, t)P(A, t)$$

Lo que indica que la probabilidad de que el sistema permanezca en A en $t + dt$ corresponde a:

$$P(A, t + dt) = P(A, t) - P_r(t, t + dt|A, t)P(A, t)$$

Si reordenamos la expresión obtenemos:

$$P(A, t + dt) - P(A, t) = -P_r(t, t + dt|A, t)P(A, t)$$

Si ahora tomamos el límite cuando $dt \rightarrow 0$

$$\lim_{dt \rightarrow 0} \frac{P(A, t + dt) - P(A, t)}{dt} = \lim_{dt \rightarrow 0} \frac{-P_r(t, t + dt|A, t)P(A, t)}{dt}$$

La parte izquierda de la ecuación corresponde exactamente a la definición de un diferencial de la probabilidad de A respecto al tiempo de forma que:

$$\lim_{dt \rightarrow 0} \frac{P(A, t + dt) - P(A, t)}{dt} = \frac{dP(A, t)}{dt}$$

La parte derecha de la ecuación corresponde a la tasa de cambio instantáneo (tasa de transición por unidad de tiempo) multiplicado por la probabilidad $P(A, t)$, por lo que:

$$\lim_{dt \rightarrow 0} \frac{-P_r(t, t + dt|A, t)P(A, t)}{dt} = -r P(A, t)$$

De tal forma que:

$$\frac{dP(A, t)}{dt} = -r P(A, t)$$

Ahora podemos integrar con respecto al tiempo considerando que nuestro estado inicial $P(A, t)=1$, por lo que:

$$P(A, t) = e^{-rt}$$

Ecuación 2-9.

Si ahora calculamos la probabilidad para B ,

$$P(B, t) = 1 - P(A, t) = F(t) = 1 - e^{-rt}$$

Ecuación 2-10

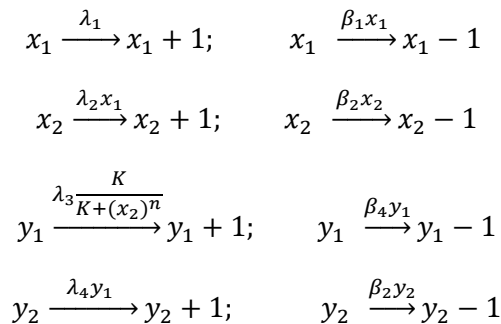
Lo que corresponde exactamente con la distribución de tiempos de reacción que postulamos para el algoritmo de Gillespie. Por tanto, Gillespie es una reformulación numérica de la CME, independientemente de las condiciones de reacción o la heterogeneidad del medio.

Código MATLAB para una trayectoria

En el caso concreto de la regulación por TetR, nuestro sistema de estudio se compone de cuatro componentes: el mRNA de TetR, TetR, el mRNA de GFP, producida desde el promotor p_{tetA} , y GFP. Denominaremos x a los componentes de TetR (x_1 al mRNA y x_2 a la proteína) mientras que el mRNA y GFP producidos a partir de p_{tetA} los denominaremos y_1 e y_2 . Asumimos que tanto los mRNAs como las proteínas se degradan de acuerdo a una cinética de primer orden βx_i , donde β es la tasa de degradación. En general, salvo para aquellas variantes en las que se indique lo contrario, $\beta_{mRNA} \gg \beta_{prot}$. Asumimos también que la función de regulación (función de transferencia) entre TetR y su promotor regulado puede ser aproximada por una función de tipo Hill con dos grados de libertad correspondientes a k , la

constante de represión y n , el grado de no-linealidad (cooperatividad aparente) del sistema.

Bajo estas asunciones, las CMEs que describen el sistema completo son:



Ecuación 2-11. Descripción de las reacciones elementales en el sistema regulador de TetR/ p_{tetA} .

Para la simulación de trayectorias comenzamos con $t = 0$ y $x_i^{t=0} = y_i^{t=0} = 0$. Dejamos al algoritmo iterar hasta alcanzar el estado estacionario. En Matlab el código correspondiente para estas simulaciones es:

CÓDIGO #1: Simulación para una trayectoria.

```
% Code for single-cell trajectories using Gillespie Algorithm
% System of study: Transcriptional repressor acting on a single target
% promoter.
%-----
%STATE VARIABLES
%r= TetR mRNA
%T=TetR protein
%R= GFP RNA
%P= GFP

% REACTIONS

% r---> r+1 rate = a0 TetR mRNA synthesis
% T----> T+1 rate= a1*r TetR protein synthesis
%R---->R+1 rate=a2*(K/((T*))+K)(K/((T*))+K)) RNA synthesis
%P---->P+1 rate=a3*R Protein synthesis

% r----->r-1 rate=b0*r TetR mRNA degradation
% T-----> T-1 rate =b1*T TetR protein degradation
%R----->R-1 T= R*b2 RNA degradation
%P-----> P-1 T=P*b3 Protein degradation
%-----
clear all

%Initialize cell concentrations
t=0; %initial time
r=0; % initial TetR mRNA
T=0; % initial TetR protein
R=0; % initial GFP mRNA
P=0; % initial GFP protein
% Propensities %calibrated in minutes so 1 mRNA degradation per 3 min.
a0=1;% TetR mRNA synthesis rate;
a1=1; % TetR protein synthesis rate
a2=4; % GFP mRNA synthesis rate
a3=2 ; % GFP protein synthesis rate
```

```

ab=0.04;% background transcription rate

b0=0.3; % TetR mRNA degradation rate
b1=0.005; % TetR protein degradation rate
b2=0.3; % GFP mRNA degradation rate
b3=0.005; % GFP protein degradation rate

K=100; % TetR repression constant

% Iterations, time and statistic vectors
sampler=1E-3;

N=20000;
h=zeros((N+1),5);
tau=0;
ii=1;
step=sampler;
h(ii,1:5)=[t r T R P ];

for t=1:N % N iterations

% rates vector
    rates= [ (a0) (a1*r) (ab+(a2*(K^4/((K^4)+(T^4)))) (a3*R) (b0*r) (b1*T)
(b2*R) (b3*P)];
    kT=sum(rates);

% Record time statistics
u_time=rand; % Pick random numbers
u_reaction=rand;
tau=-1/kT*log(1-u_time); % Update time statistics
t=t+tau;
    ii=ii+1;
    h(ii,1:5)=[t r T R P ];
    q(ii,1:8)=[rates];

av=0; % Gillespie pointers
j=1;

while av<kT*u_reaction;
    av=av+rates(j);
    j=j+1;
end

    switch j-1
        case 1
            r=r+1;
        case 2
            T=T+1;
        case 3
            R=R+1;
        case 4
            P=P+1;
        case 5
            r=r-1;
        case 6
            T=T-1;
        case 7
            R=R-1;
    end

```

```

case 8
    P=P-1;
otherwise
    disp('error')
end

end

results=h;

```

Código MATLAB para poblaciones y condiciones

Para obtener resultados poblacionales, el código anterior se itera para un total de N células. Para cada célula tomamos sus valores finales de tiempo, mRNA y proteínas y se guardan en una matriz que posteriormente utilizaremos para elaborar las distribuciones. Para asegurarnos de que el muestreo se realiza en estado estacionario, se traza la trayectoria de cada célula durante un total de 20000 iteraciones.

Además, las N células son simuladas en condiciones diferentes. Estas condiciones representan distintos niveles de inducción de arabinosa, lo que se traduce en diferentes niveles de producción del mRNA de *tetR* (λ_1). El código considera un nivel inicial y un nivel final de inducción y lo divide en nn intervalos logarítmicamente distribuidos.

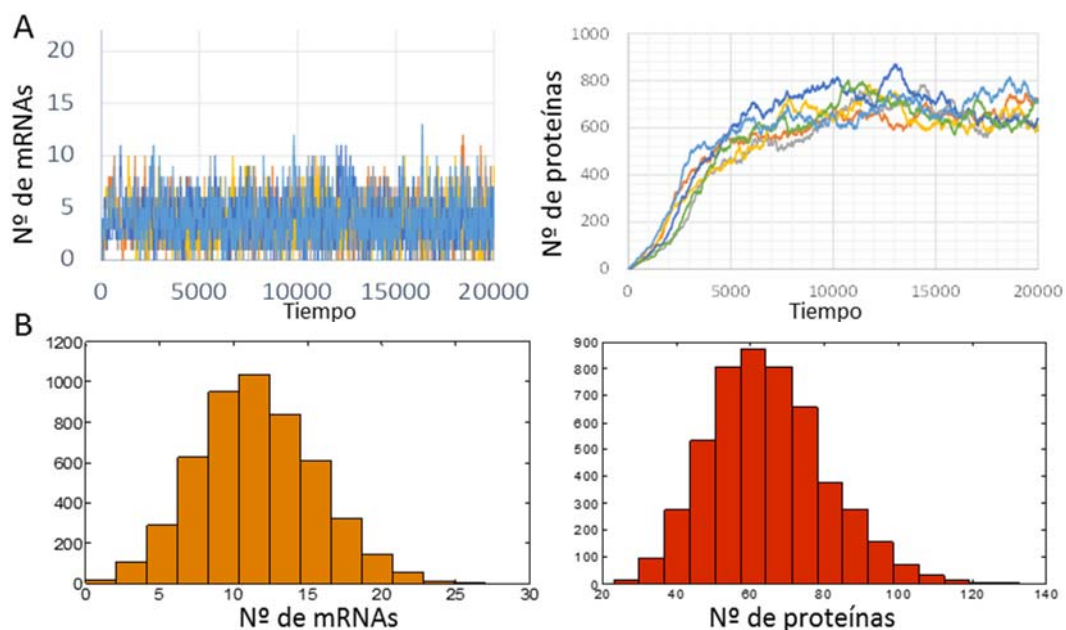


Figura 2-11. (A) Trayectorias de la CME de N células para mRNA (panel izquierdo) y para proteína (panel de la derecha). Cada trayectoria es única debido a la estocasticidad inherente a los procesos de transcripción-traducción. No obstante, si se deja evolucionar al sistema hasta el estado estacionario, se obtiene una distribución invariante en el tiempo para la cantidad de mRNAs y proteínas. (B) Distribuciones en estado estacionario para las trayectorias representadas en A. Aunque el valor individual de cada célula es variable, la distribución poblacional para un sistema dado en estado estacionario es constante.

CÓDIGO #2: Análisis de poblaciones

```
% Gillespie model for N cells and nn induction levels of TetR.
%Model
%r= TetR mRNA
%T=TetR protein
%R= GFP RNA
%P= GFP

% REACTIONS

% r---> r+1 rate = a0 TetR mRNA synthesis
% T----> T+1 rate= a1*r TetR protein synthesis
%R---->R+1 rate=a2*(K/((T*))+K)(K/((T*))+K)) RNA synthesis
%P---->P+1 rate=a3*R Protein synthesis
% r----->r-1 rate=b0*r TetR mRNA degradation
% T-----> T-1 rate =b1*T TetR protein degradation
%R----->R-1 T= R*b2 RNA degradation
% P-----> P-1 T=P*b3 Protein degradation

%-----
clear all

nn=7; % number of induction levels simulated
cc=500;% number of cells per induction level
xx=zeros((cc+1),6);%results per induction array
results=zeros(1,6);%total cell results
A=0.1;
B=10;
LA = log10(A); LB = log10(B); %Choose the min (A) and the max TetR mRNA
production rate
RANGE=LB-LA; % Calculate the range of induction levels
INLEVEL=RANGE/nn; % Calculate the step for the induction levels tested

for indd=1:nn %number of induction levels
    %Initialize cell concentrations
    t=0; %initial time
    r=0; % initial TetR mRNA
    T=0; % initial TetR protein
    R=0; % initial GFP mRNA
    P=0; % initial GFP protein

    for qq=1:cc;

        % Propensities %calibrated in minutes so 1 mRNA degradation per 3 min.
        a0= 10^LA;% TetR mRNA synthesis rate;
        a1=1; % TetR protein syntehsis rate
        a2=4; % GFP mRNA syntehsis rate
        a3=2 ; % GFP protein synthesis rate
        ab=0.04;% background transcription rate

        b0=0.3; % TetR mRNA degradation rate
        b1=0.005; % TetR protein degradation rate
        b2=0.3; % GFP mRNA degradation rate
        b3=0.005; % GFP protein degradation rate

        K=100; % TetR repression constant

        % Iterations, time and statistic vectors
```

```

sampler=1E-3;

N=20000;
h=zeros((N+1),5);
tau=0;
ii=1;
step=sampler;
h(ii,1:5)=[t r T R P ];

for t=1:N % N iterations

% rates vector
rates= [ (a0) (a1*r) (ab+(a2*(K^4/((K^4)+(T^4)))) (a3*R) (b0*r) (b1*T)
(b2*R) (b3*P)];
kT=sum(rates);

% Record time statistics
u_time=rand; % Pick random numbers
u_reaction=rand;
tau=-1/kT*log(1-u_time); % Update time statistics
t=t+tau;

ii=ii+1;
h(ii,1:5)=[t r T R P ];
q(ii,1:8)=[rates];

av=0; % Gillespie pointers
j=1;

while av<kT*u_reaction;
    av=av+rates(j);
    j=j+1;
end

switch j-1
case 1
    r=r+1;

case 2
    T=T+1;

case 3
    R=R+1;

case 4
    P=P+1;

case 5
    r=r-1;

case 6
    T=T-1;

case 7
    R=R-1;

case 8
    P=P-1;

```

```
        otherwise
            disp('error')

    end

end

xx(qq,1:6)=[a0 t r T R P ];

end
results=vertcat(results,xx);
LA=LA+INLEVEL; %Update TetR induction level
end
```

2.9.3. Medida de la información mutua (I):

De acuerdo con la definición de Shannon, la información mutua (I) entre dos variables (X, Y) corresponde a:

$$I(X; Y) = \int_{XYmin}^{XYmax} dP(X, Y) \log \frac{dP(X, Y)}{d(P(X)P(Y))}$$

Ecuación 2-12. Información mutua.

I es una función no-lineal de una distribución de probabilidad conjunta $P(X, Y)$ y dos distribuciones marginales $P(X)$ y $P(Y)$. $P(X)$ y $P(Y)$ forman dos espacios mensurables, por lo que para determinar $I(X; Y)$ necesitamos determinar los valores que una de las variables va adquiriendo a medida que cambia la otra. Esto plantea ciertos problemas a la hora de determinar experimentalmente I :

- I está condicionada por las distribuciones marginales sobre las que se realiza la medida. Esto es, para determinar la información mutua entre un regulador transcripcional (X) y su promotor regulado (Y), necesitamos medir Y frente a todo el rango de concentraciones de X que *la célula es capaz de experimentar*. A esto lo denominamos **problema del espacio significativo**.
- Los datos que obtenemos experimentalmente se refieren a X e Y , pero no a sus probabilidades. Estas últimas deben inferirse de acuerdo al conjunto de los datos obtenidos, lo que implica realizar una partición del espacio mensurable en ambas dimensiones. En otras palabras, asignar una probabilidad a un valor de X implica determinar el rango posible de X (lo que remite al problema anterior), pero también determinar el número de particiones (*bins*) de X para las que podemos asignar una probabilidad en base a las frecuencias observadas. Imaginemos por ejemplo una variable aleatoria X que toma valores entre 0 y 100, de la cual realizamos 3 medidas y

obtenemos $X_1=54$, $X_2=15$ y $X_3=60$. Obviamente $P(X \leq 10)$ es distinta de $P(X \leq 25)$, por lo que obtendremos distribuciones de probabilidad diferentes dependiendo de si dividimos X en 10 o en 4 categorías discretas. A este problema lo llamaremos **problema categorial**.

Problema del espacio significativo

Medir la información mutua entre las concentraciones de un regulador transcripcional (X) y su promotor regulado (Y) implica determinar el estado de una variable en función de todas las configuraciones posibles de la otra. En nuestro caso concreto, medir la respuesta del promotor p_{tetA} ante todas las concentraciones fisiológicamente relevantes de regulador TetR. Para ello, TetR:mKate2 fue inducido a distintos niveles desde el promotor p_{BAD} . Como nivel basal se emplearon condiciones en las que la expresión de p_{BAD} se encuentra totalmente reprimida (ausencia de arabinosa y presencia de glucosa, para evitar activación por CRP). Para obtener máximos niveles de TetR se indujo el promotor p_{BAD} mediante arabinosa y en presencia de glicerol como única fuente de carbono. Para comprobar que los niveles máximos obtenidos por este método eran superiores a los producidos por el promotor nativo (p_{tetR}), se compararon los niveles producidos por nuestras construcciones frente a los alcanzados por TetR bajo su promotor nativo (Figura 3-29, Resultados).

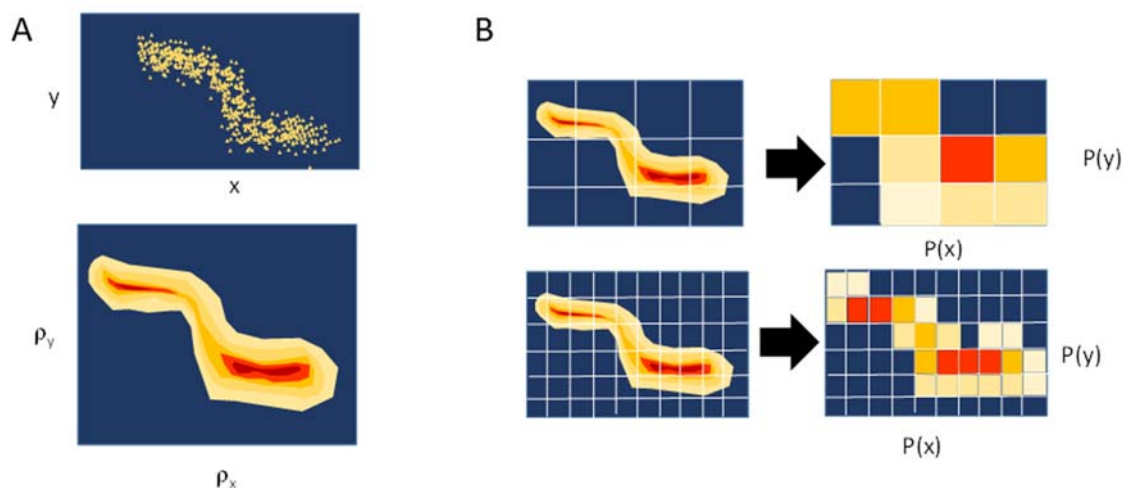


Figura 2-12. (A) Mapa *input/output*. A partir de un número de observaciones (panel superior) de pares x, y se puede inferir un mapa de densidad para ambas variables (panel inferior). Sin embargo, para convertir densidades en probabilidades es necesario realizar una partición del espacio. (B) Partición del espacio mensurable mediante el método de tamizado. El mapa de observaciones (paneles de la derecha) se subdivide en una matriz de categorías de tamaño decreciente. A cada categoría (*bin*) se le puede asignar entonces una probabilidad dada a partir de la frecuencia de observaciones. Una matriz con categorías muy grandes (paneles superiores) produce una pérdida de resolución y una infraestimación de la información mutua. Por el contrario, una matriz con categorías muy pequeñas (paneles inferiores) requiere de un número más alto de observaciones para poder asignar probabilidades de forma adecuada.

Problema categorial

Calcular I a partir de datos experimentales requiere transformar la nube de observaciones de niveles de *input/output* (x,y) en distribuciones de probabilidad $P(x,y)$. Para ello, es inevitable dividir un espacio continuo en categorías (*bins*) a los que poder asignar probabilidades. La partición del espacio mensurable no es un problema trivial: dividir el espacio en un número insuficiente de *bins* conlleva una pérdida de resolución (Figura 2-12), puesto que cada *bin* contiene distintos niveles (x,y) a los cuales se declara equiprobables. Sin embargo, aumentar el número de *bins* indefinidamente tampoco resuelve este problema: a medida que se aumenta el número de *bins* se necesita un muestreo mayor para poder asignar $P(x,y)$ de manera robusta. Este problema es análogo al de la sensibilidad de una cámara fotográfica: aumentando la cantidad de píxeles se consigue mayor resolución, pero disminuye la sensibilidad y se requieren exposiciones mayores. Existen distintos métodos para calcular el tamaño óptimo de *bin* en función de las propiedades estadísticas de las distribuciones marginales. Algunos de estos métodos asumen distribuciones normales (regla de Scott) o sólo tienen en cuenta el número de observaciones, sin importar el rango (regla de Sturges). La regla de Scott parece poco apropiada para datos de expresión genética (que tienden a mostrar distribuciones Poissonianas) y el rango de nuestros datos es demasiado disperso (varios logs de diferencia) para aplicar la regla de Sturges. El método de Freedman-Diaconis ofrece otra alternativa en la que los rangos intercuartiles del 25 y el 75 % son utilizados para determinar el tamaño óptimo del *bin*, independientemente de la distribución subyacente. Sin embargo, Freedman-Diaconis puede dar lugar a distintos números óptimos de *bin* para variables diferentes. Dado que la computación de la información mutua requiere que las dos variables consideradas sean discretizadas de manera idéntica, Freedman-Diaconis tampoco ofrece garantías de rendir el tamaño óptimo del *bin*. Para evitar estos problemas, utilizaremos un método análogo al de la inferencia estadística abstracta desarrollado por Ulf Grenander (Grenander, 1981). El método de Grenander, o método de tamizado, utiliza una propiedad fundamental de la información mutua (la desigualdad en el procesamiento de información) que establece que, siendo S y T dos funciones arbitrarias de dos variables aleatorias x e y :

$$I(x;y) \geq I(S(x);T(y))$$

Ecuación 2-13.

Esto es, aplicando cualquier partición arbitraria sobre nuestras variables de estudio podemos obtener el límite inferior de información transmitida. Si realizamos un número sistemático de particiones, aquella con un valor de I máximo será la más cercana al valor verdadero de $I(x;y)$. En la práctica, este método de tamizado consiste en tomar el valor máximo y el

mínimo de la variable con mayor rango de datos, y dividir el espacio en un número n de *bins* equidistantes. Iteramos este proceso aumentando el número de *bins* y computamos la información mutua para cada partición. Para ello utilizamos la ecuación de Shannon para valores discretos que establece la relación entre la información mutua y la entropía (H):

$$I(x; y) = H(x) + H(y) - H(x, y)$$

Ecuación 1-6

Donde H equivale a:

$$H(x) = - \sum_i P(x_i) \log_2(P(x_i))$$

Ecuación 1-4. Entropía de Shannon.

Tomando logaritmos en base 2 para utilizar el bit como unidad de información. El único problema que presenta esta aproximación es que, dado que la entropía no puede ser negativa, un aumento del número de *bins* conlleva automáticamente un aumento en H (Paninski, 2003). Esta inflación puede eliminarse corrigiendo la estimación de la entropía por el número de *bins* utilizados. Para n *bins* y N muestras:

$$E(I) = \frac{(n-1)^2}{2N \log_2(2)}$$

Ecuación 2-14. Error cometido por el tamizado.

El siguiente código de Matlab se utilizó para calcular $I(x; y)$ utilizando el método de tamizado de Grenander, con corrección del error inducido por el número de *bins*.

Código MATLAB para el cálculo de $I(x;y)$

```
% Script for the calculation of Entropy and Mutual Info for x,y pairs under
% variable binning
```

```
% Input: introduce two vectors, x and y. Assuming column vectors
l=length(x);
xx=x';
yy=y';
%m=[xx yy]; % Uncomment when introducing column vectors
m=[x y]; % Uncomment when introducing row vectors
```

```
% Calculating boundaries for binning
minx= min(x); %minimal value of x
maxx= max(x); %maximal value of x
miny= min(y); %minimal value of y
maxy= max(y); % maximal value of x
```

```
n=10; % initial bin number
nn=100; %final bin number
for ii=1:nn
```

```
% Binning

minBinX=log10(minx);
maxBinX=log10(maxx);
minBinY=log10(miny);
maxBinY=log10(maxy);
BinX=logspace(minBinX,maxBinX,n);
BinY=logspace(minBinY,maxBinY,n);

% Uncomment for linear binning
%BinX=linspace(minx,maxx,n);
%BinY=linspace(miny,maxy,n);

% Bin marginal distributions and compute Entropy
    % Get bin counts
    hx=histc(x,BinX);
    hy=histc(y,BinY);
    % Compute Probabilities
    Px=hx./sum(hx);
    Py=hy./sum(hy);
    % Compute logarithms and products
    logPx=log2(Px+eps); % we sum eps to avoid log(0)
    ProductX=Px.*logPx;
    logPy=log2(Py+eps); %we sum eps to avoid log(0)
    ProductY=Py.*logPy;
    % Compute H
    Hx=-1.*sum(ProductX);
    Hy=-1.*sum(ProductY);

    % Computing P(xy)
    edges = {(BinX),(BinY)};
    hxy=hist3(m,'Edges',edges);
    Pxy=hxy/sum(hxy(:));
    % Log , products and Entropy
    LogPxy=log2(Pxy+eps);
    ProductXY=Pxy.*LogPxy;
    Hxy=-1.*sum(ProductXY(:));

% Computing MUTUAL INFO

I=Hx+Hy-Hxy;

%Binning error
err=((n-1)^2)./(2*1*log2(2));
II=I-err;
r(ii,1:2)=[n,II];
n=n+1;
end
```

Resultados

3. Resultados

Con el objetivo principal de medir la regulación transcripcional de una forma precisa y sencilla, se realizaron una serie de experimentos fundamentados en la cuantificación de fluorescencia producida por proteínas reporteras. Para un mejor seguimiento, podemos dividir los resultados experimentales en dos bloques: cuantificación en cultivo por fluorímetro y cuantificación en célula única mediante microscopio de epifluorescencia.

3.1. Cuantificación de la regulación transcripcional en cultivo mediante fluorímetro

Siguiendo la metodología descrita en el apartado 2.7 se analizó la actividad transcripcional de un promotor en presencia de distintos niveles de su regulador, para cuantificar dicho mecanismo de regulación. En este procedimiento, el promotor objeto de estudio origina la síntesis de proteínas fluorescentes que, a su vez, generan la fluorescencia que es medida y permite cuantificar la actividad promotora. El método propuesto realiza el análisis sobre cultivos en crecimiento y por lo tanto, depende de la capacidad técnica de medir fluorescencia y de discernir entre los distintos estados del sistema, al mismo tiempo que se realiza la incubación de los cultivos bacterianos.

3.1.1. Actividad transcripcional del promotor

Como primera aproximación se midió la autofluorescencia de la cepa BW27783 frente a la fluorescencia producida por el promotor diana p_{tetA} y otros promotores, que sirvieron para contextualizar el nivel de actividad: p_R del fago λ , de amplio uso en biología sintética; p_{Ant} del plásmido R388, de interés en nuestro laboratorio; y BBa_J23101, un promotor sintético propuesto como estándar (Kelly *et al.*, 2009). Las bacterias se incubaron en el propio fluorímetro a 37 °C en medio M9 con los antibióticos apropiados para cada cepa. Se realizaron medidas tanto de fluorescencia como de absorbancia, por lo que se pudo medir la actividad y cuantificar el crecimiento poblacional al mismo tiempo.

Utilizando los valores obtenidos durante el crecimiento exponencial, se calculó la actividad del promotor y el tiempo de generación (Figura 3-1). Se observó que el tiempo de generación para BW27783 era de 60 ± 4 min y su autofluorescencia se utilizó como valor de fondo. Las cepas portadoras de los diferentes vectores reporteros incrementaron hasta en 8 min su tiempo de generación. El promotor p_{Ant} demostró ser el más activo, seguido de p_{tetA} , p_R y BBa_J23101, en

orden decreciente. El nivel de producción de proteína fluorescente no presentó una relación directa con el descenso en la tasa de crecimiento.

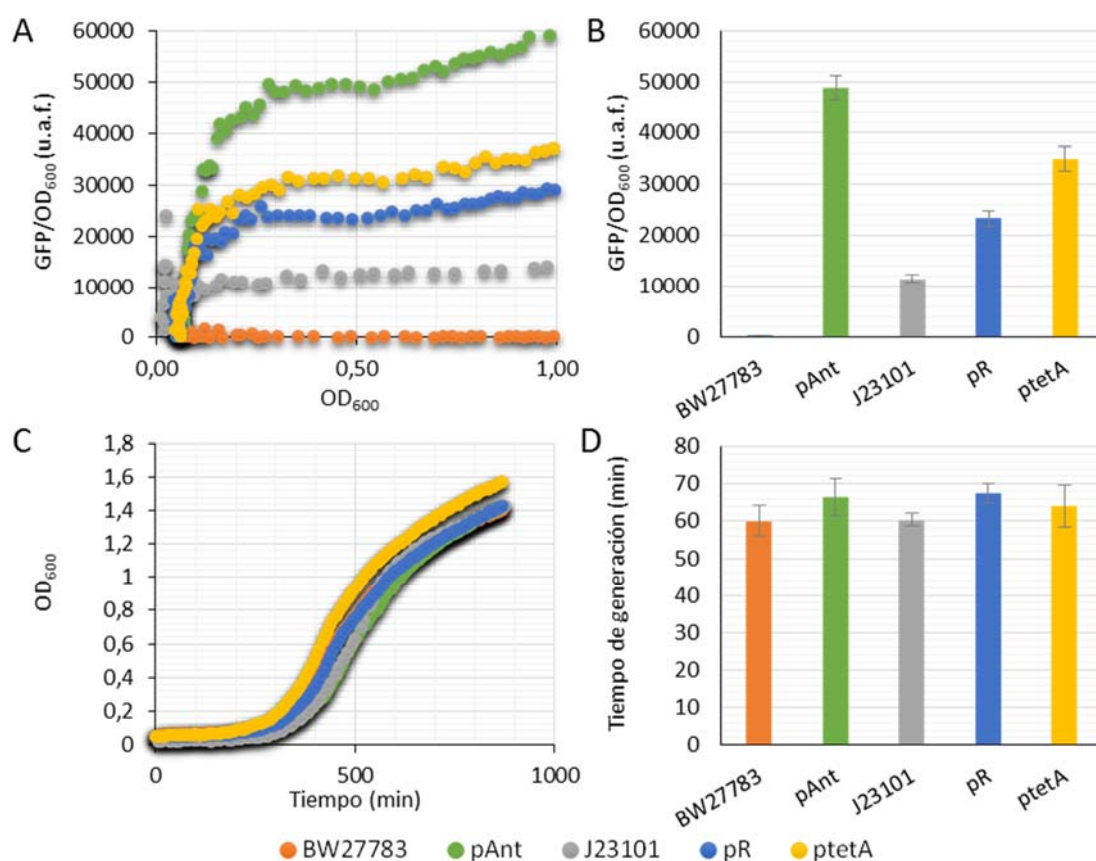


Figura 3-1. Cuantificación de la actividad de los promotores. La cepa BW27783, con los plásmidos reporteros para los promotores correspondientes al código de colores, se analizó en el fluorímetro como se detalla en el apartado 2.7 de Materiales y Métodos (A) Fluorescencia producida (GFP/OD₆₀₀) frente a los valores de crecimiento del cultivo (OD₆₀₀) (B) Valores medios de la actividad promotora durante el crecimiento exponencial del cultivo. Las barras de error reflejan la desviación estándar de las distintas medidas. (C) Crecimiento de los cultivos representado como OD₆₀₀ frente al tiempo en min. (D) Tiempos de generación en min y su desviación estándar.

El promotor a estudio en este trabajo, p_{tetA}, presentó una actividad de 34994±2358 u.a.f. La cepa con el vector pRRG13, reportero para p_{tetA}, tuvo un tiempo de generación 64±6 min.

3.1.2. Caracterización de los sistemas de expresión

Como se explicó en el apartado metodológico, se pudo expresar el regulador de manera gradual para analizar su actividad en su promotor diana. Para ello, se utilizó el sistema de expresión del operón arabinosa incluido en dos plásmidos: uno con origen de replicación p15A de unas 18 copias por cromosoma (Chang and Cohen, 1978); y otro con origen de replicación RK2 y unas 4-7 copias por cromosoma (Figurski *et al.*, 1979; Grinter, 1984). No obstante, y previamente a su uso con TetR, como control para caracterizar el comportamiento de los sistemas de inducción, se construyeron dos vectores de expresión de GFP: pRRG57 con el origen RK2 y pRRG58 con el origen p15A. Dos cepas, una con cada plásmido, se cultivaron en el

fluorímetro a 37 °C en M9. Como fuente de carbono se utilizaron, alternativamente, glucosa por un lado o glicerol por otro. Para ambos medios, se estableció un gradiente de arabinosa desde 0 hasta 16 μ M. Los datos de fluorescencia obtenidos (Figura 3-2) confirmaron la inducción progresiva de p_{BAD} con el incremento de arabinosa en el medio. El plásmido con mayor número de copias, pRRG58, generó mayor fluorescencia que pRRG57 en todas las concentraciones de arabinosa, hasta que los valores inferiores rebasaron el límite de sensibilidad de la técnica. El efecto de la glucosa (Figura 3-2.B) frente al glicerol (Figura 3-2.A) quedó también contrastado. La incubación en M9 con glucosa disminuyó drásticamente la actividad promotora. Si bien las concentraciones inferiores de arabinosa utilizadas registraron valores de autofluorescencia, en concentraciones superiores de arabinosa la diferencia glicerol/glucosa fue notable. Por ejemplo, con arabinosa 16 μ M, la cepa portando pRRG58 generó 99425 ± 6268 u.a.f. en el cultivo con glicerol, mientras que en presencia de glucosa el valor fue de 1745 ± 322 . Los datos de crecimiento de los cultivos también fueron analizados (Figura 3-2.C), observándose un incremento en la velocidad de crecimiento de los cultivos, como cabría esperar, en los medios con glucosa. No obstante, no se detectó cambio en el tiempo de generación relacionado con el gradiente de arabinosa. Estos resultados servirían como referencia para estimar la cantidad de TetR producida en las distintas condiciones de inducción en los ensayos posteriores.

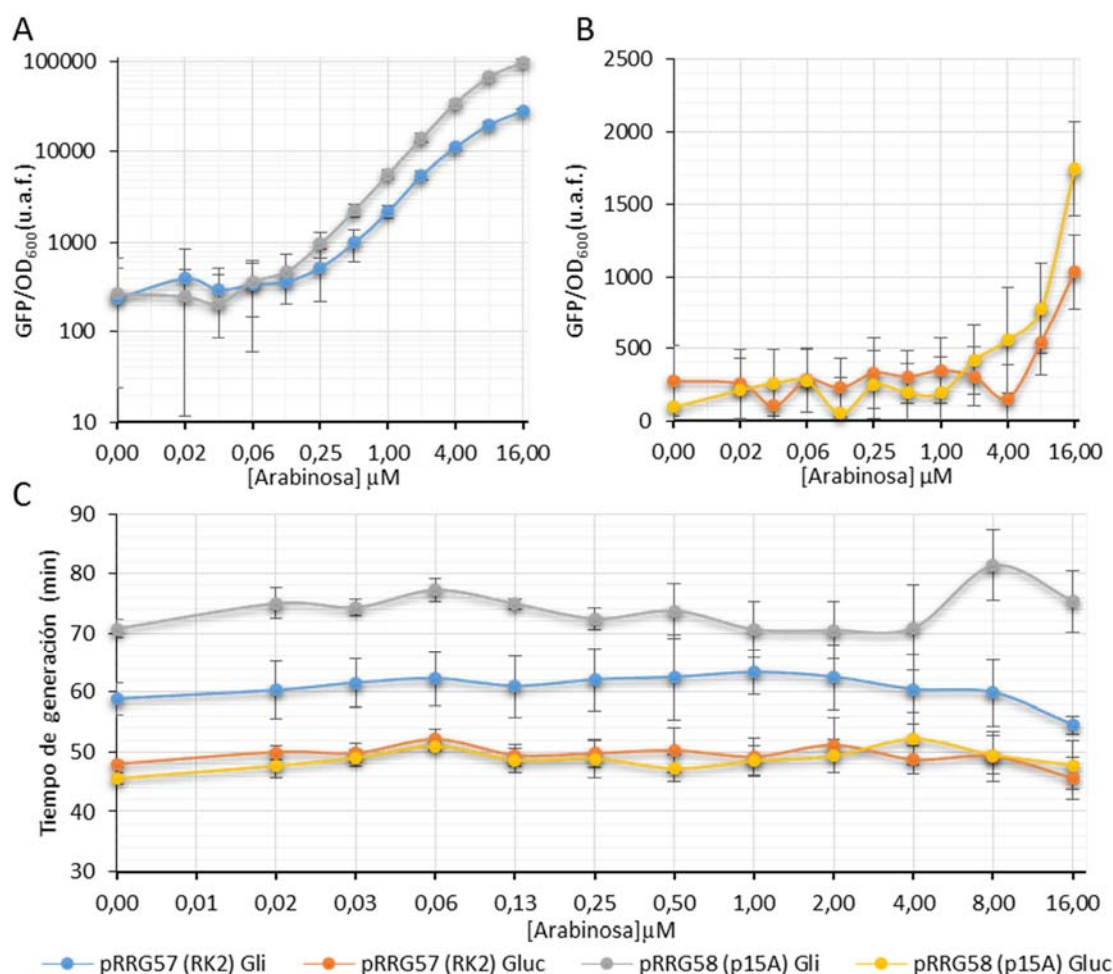


Figura 3-2. Actividad de p_{BAD} en los vectores pRRG57 y pRRG58. Los cultivos se analizaron en el fluorímetro como se detalla en el apartado 2.7 de Materiales y Métodos. Se estudiaron diferentes condiciones inductoras del promotor: (A) crecimiento en M9 con glicerol y un gradiente de arabinosa de 0 a 16 μM ; (B) crecimiento en M9 con glucosa y el mismo gradiente de arabinosa. Se representa la actividad promotora relativizada por el crecimiento ($\text{GFP}/\text{OD}_{600}$) frente al gradiente de arabinosa. (C) Tiempos de generación de los cultivos. El código de colores en la gráfica relaciona plásmido y medio de cultivo. En los tres paneles las barras de error representan la desviación estándar de los distintos cultivos analizados y el gradiente de arabinosa se representa en escala logarítmica en base 2.

3.1.3. Actividad reguladora de TetR

El siguiente paso fue el análisis del efecto de TetR sobre su promotor diana. La cepa portadora de pRRG13 se transformó con dos vectores de expresión de TetR, el vector pRRG1 (p15A) y pRRG9 (RK2) (Figura 3-3). Se crecieron los cultivos en el fluorímetro en medio M9 suplementado con glicerol y se añadieron los antibióticos correspondientes a las resistencias acarreadas por los vectores. Se realizó un gradiente de arabinosa en diluciones seriadas de 1/2 desde 16 hasta 0 μM , induciendo diferentes niveles de producción de TetR.

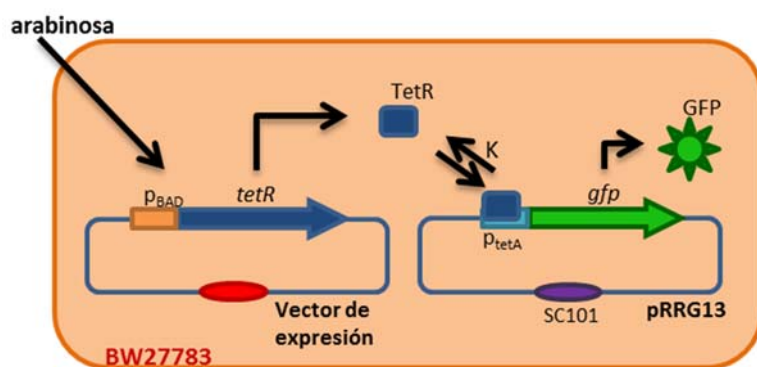


Figura 3-3. Configuración de las cepas experimentales con dos vectores. El vector de expresión produce el represor TetR como respuesta a la presencia de arabinosa. A su vez, TetR controla negativamente la expresión del gen *gfp*, codificado en el vector reportero pRRG13, mediante su unión al promotor *p_{tetA}*. El vector de expresión puede tener origen de replicación p15A (pRRG1) o RK2 (pRRG9) (representado en la figura por un óvalo rojo).

Los datos obtenidos se presentan en la Figura 3-4. La velocidad de crecimiento de la cepa se mantuvo constante para ambas construcciones a medida que se incrementaba la concentración de arabinosa, hasta llegar a 4 μM donde se detectó una notable disminución. La cepa con pRRG1 incrementó su tiempo de generación, cercano a 70 min, hasta 403 ± 23 min en presencia de 16 μM de arabinosa. La cepa con pRRG9, por su parte, cambió de una media de 77 a 142 ± 14 min. La fluorimetría, igualmente, mostró un nivel constante hasta que, simultáneo al cambio del tiempo de generación de las cepas, se produjo un aumento de la fluorescencia. La sola presencia de TetR redujo drásticamente el nivel de actividad del promotor *p_{tetA}*, dando valores cercanos a 0 con pRRG1 y en torno a 900 u.a.f. con pRRG9. Por tanto, el represor TetR se comportó como tal, reprimiendo la actividad del promotor hasta obtener niveles de fondo (la autofluorescencia de la cepa). El aumento de TetR no produjo un mayor efecto represor detectable mediante esta técnica. En cambio, se observó un incremento de la fluorescencia en las mayores inducciones de TetR, explicadas por la menor tasa de crecimiento y, por tanto, una mayor acumulación intracelular de proteína fluorescente, y no por una mayor actividad promotora. Estos fenómenos indicaron que altas concentraciones de TetR son dañinas para la célula.

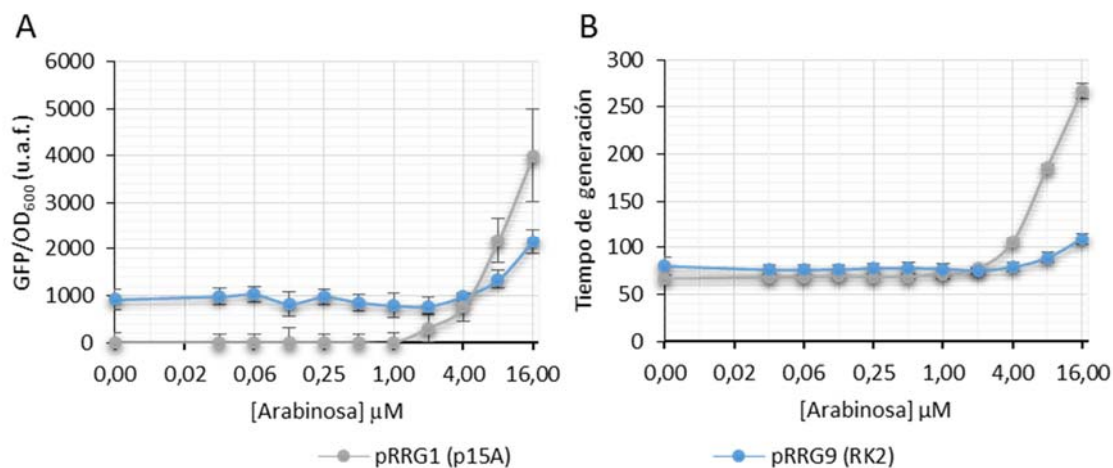


Figura 3-4. Actividad del promotor p_{tetA} en presencia de pRRG1 (gris) y pRRG9 (azul). Los cultivos crecieron en M9 con glicerol y se analizaron en el fluorímetro como se detalla en el apartado 2.7 de Materiales y Métodos. (A) Se representa la actividad promotora relativizada por el crecimiento (GFP/OD_{600}) frente al gradiente de arabinosa. (B) Tiempos de generación de los cultivos. En ambos paneles las barras de error representan la desviación estándar y el gradiente de arabinosa se representa en escala logarítmica en base 2. Los orígenes de replicación de los vectores de expresión se indican entre paréntesis en la leyenda de las gráficas.

Con esta configuración experimental no era posible conseguir concentraciones inferiores de represor, y se optó por cambiar el glicerol por la glucosa como fuente de carbono.

Se analizó la cepa BW27783 con los vectores pRRG9 y pRRG13 (Figura 1-5). Se cultivó en el fluorímetro a 37 °C en medio M9 suplementado con glucosa y los antibióticos correspondientes. El nivel de actividad registrado sin inducir la síntesis de represor se situó en 20701 ± 2512 u.a.f., lo que supuso un 60 % de la fluorescencia de la cepa con pRRG13 en solitario. Incrementado la concentración de arabinosa hasta 0,04 μM se observó una subida de la fluorescencia. En el incremento sucesivo de la concentración de arabinosa, la actividad de p_{tetA} fue descendiendo progresivamente, hasta alcanzar niveles de autofluorescencia. El examen de las curvas de crecimiento de la cepa en los distintos niveles de inducción demostró que las concentraciones producidas de TetR no interferían con el crecimiento. La glucosa facilitó el crecimiento celular, observándose un descenso de media hora del tiempo de generación (de ≈ 80 min en glicerol a ≈ 40 min en glucosa), pero se produjo un incremento en la variabilidad de este parámetro (desviaciones estándar de la Figura 3-5.B).

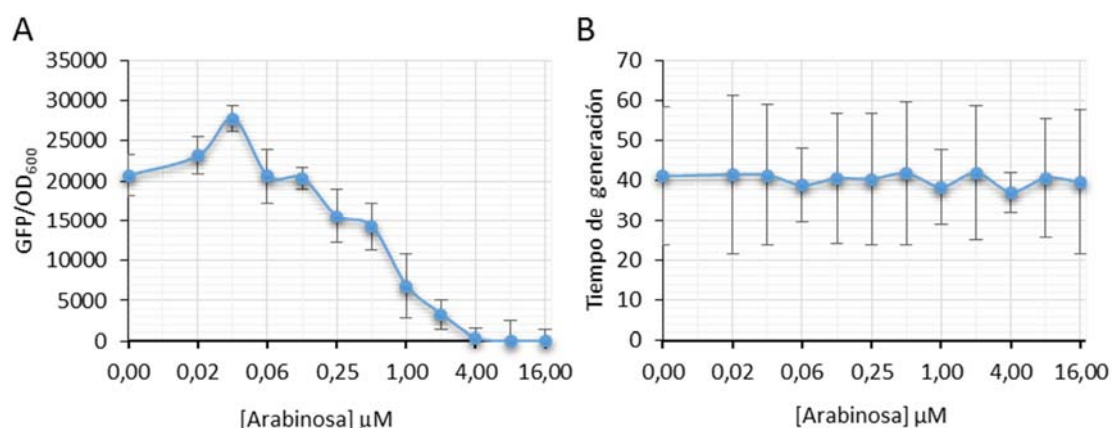


Figura 3-5. (A) Actividad del promotor p_{tetA} en presencia de pRRG9. Los cultivos crecieron en M9 con glucosa y se analizaron en el fluorímetro como se detalla en el apartado 2.7 de Materiales y Métodos. (A) Se representa la actividad promotora relativizada por el crecimiento (GFP/OD₆₀₀) frente al gradiente de arabinosa. (B) Tiempos de generación de los cultivos. En ambos paneles las barras de error representan la desviación estándar y el gradiente de arabinosa se representa en escala logarítmica en base 2.

Por lo tanto, el sistema de medida compuesto por los vectores pRRG13, pRRG9 y la cepa BW27783 permitió observar una respuesta gradual del promotor p_{tetA} frente a diferentes concentraciones de su regulador TetR. No obstante, los valores obtenidos sin inducción por arabinosa quedaron por debajo de los presentados por la cepa con sólo pRRG13, demostrando que en esas condiciones pRRG9 está produciendo una cantidad significativa de regulador. Por otra parte, en concentraciones de arabinosa superiores a 4 μ M, la fluorescencia producida por el sistema alcanzó niveles de autofluorescencia. Es decir, el rango de actividad de p_{tetA} observado con este sistema respondía a niveles de arabinosa entre 0 y 4 μ M. Volviendo a los datos obtenidos con pRRG57 (Figura 3-2.B) se comprobó que la respuesta en la actividad inductora de p_{BAD} en presencia de glucosa fue prácticamente indetectable precisamente en ese rango de concentraciones de arabinosa. Dado que se pretendía inferir la concentración de TetR producido teniendo en cuenta la actividad de p_{BAD} , nos encontramos con un problema. En resumen, el rango estimable de regulador (concentración de regulador estimada por la dinámica de inducción de p_{BAD} en pRRG57, Figura 3-2) no generó respuesta, mientras que el rango observable de actividad de p_{tetA} fue en respuesta a concentraciones no detectables de la señal.

3.1.4. Fusión TetR:mKate2

En vista de los resultados anteriores, se midió directamente tanto la entrada (TetR) como la salida (actividad de p_{tetA}) del sistema de regulación por fluorescencia. Para que la medición del regulador fuese directa se realizó una fusión traduccional de la proteína TetR con la proteína fluorescente roja mKate2 (vector pRRG62).

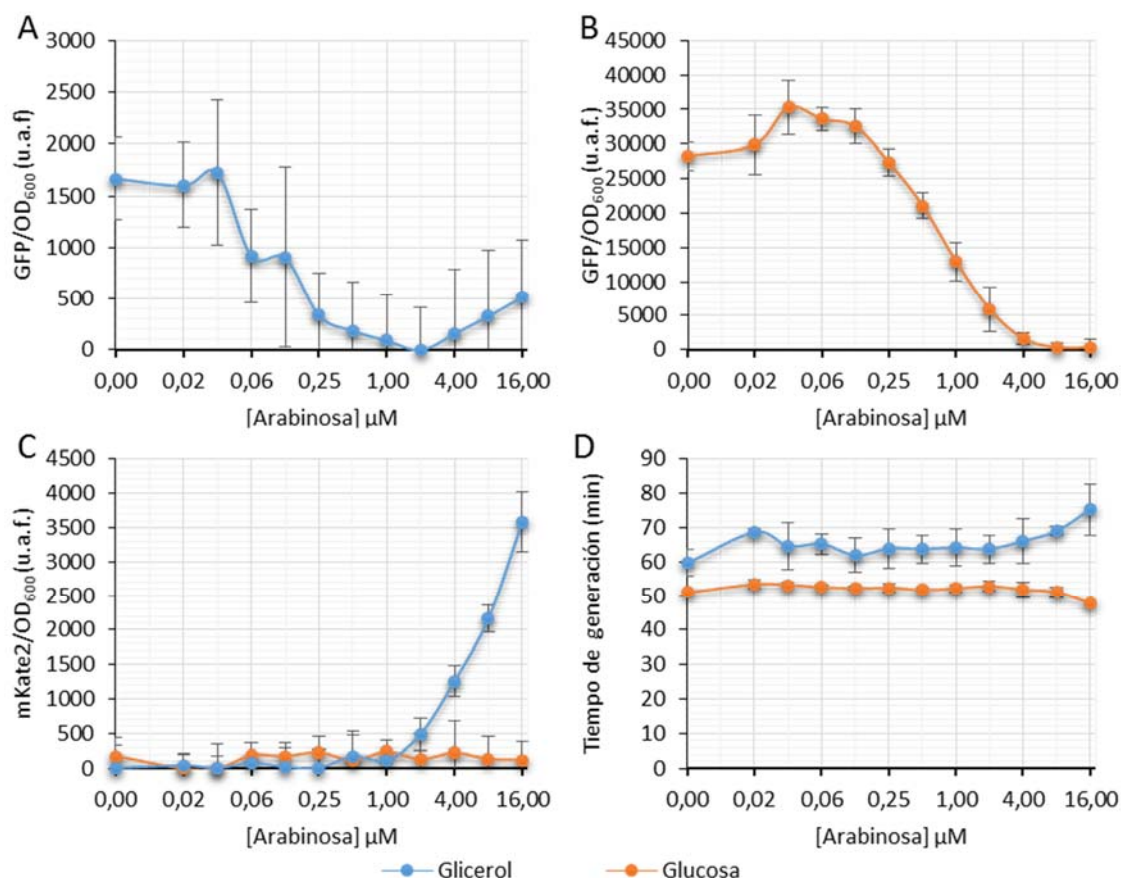


Figura 3-6. Fluorescencia roja y verde medida para BW27783 con pRRG13 y pRRG62. Los cultivos crecieron en M9 con glicerol (azul) o glucosa (naranja) y se analizaron en el fluorímetro como se detalla en el apartado 2.7 de Materiales y Métodos. Se representa la actividad promotora relativizada por el crecimiento (GFP/OD_{600}) frente al gradiente de arabinosa en cultivo: (A) en M9 con glicerol y (B) en M9 con glucosa. (C) Representación de la concentración de TetR:mKate2 producida (fluorescencia roja) relativizada por el crecimiento ($mKate2/OD_{600}$) frente al gradiente de arabinosa. (D) Tiempos de generación de los cultivos. Las barras de error representan la desviación estándar de los datos.

Al igual que en las mediciones previas, cultivos de BW27883 con los vectores pRRG13 y pRRG62 fueron incubados en el fluorímetro. Se estudiaron todas las condiciones, utilizando glucosa y glicerol como fuentes de carbono y realizando el gradiente de arabinosa 0-16 μM para ambos casos. Los resultados obtenidos (Figura 3-6.A y B) para la actividad promotora de p_{tetA} en presencia de TetR:mKate2 fueron similares a los obtenidos con TetR, aunque ligeramente más altos, indicando una disminución en la actividad represora de TetR debida probablemente a la fusión. Los tiempos de generación mostraron que, al igual que TetR, la proteína de fusión no supuso una carga para las bacterias, salvo en las concentraciones superiores de arabinosa en presencia de glicerol donde hubo cierto aumento en el tiempo de generación (Figura 3-6.D).

Por su parte, la fluorescencia medida para cuantificar TetR:mKate2 se muestra en la Figura 3-6.C. Los resultados mostraron que el rango de detección para la proteína roja, y por tanto la estima de la concentración del regulador, era similar a los obtenidos con GFP con los vectores

de expresión (Figura 3-2), siendo indetectable en las condiciones con glucosa. Representando en un gráfico los datos de actividad de p_{tetA} , medida en fluorescencia verde, frente a la concentración de TetR, medida en fluorescencia roja (Figura 3-7), se comprobó que el cambio de la señal verde no era la respuesta a un cambio en la señal roja. Solamente cuando la señal verde se encontraba en valores de fondo, la señal roja superaba el umbral de detección (≈ 100 -500 u.a.f.). Este resultado reafirmó la situación anterior: el rango detectable de regulador no generó respuesta detectable del promotor, mientras que la respuesta detectable del promotor se observó en un rango de concentraciones de regulador indetectable por esta técnica.

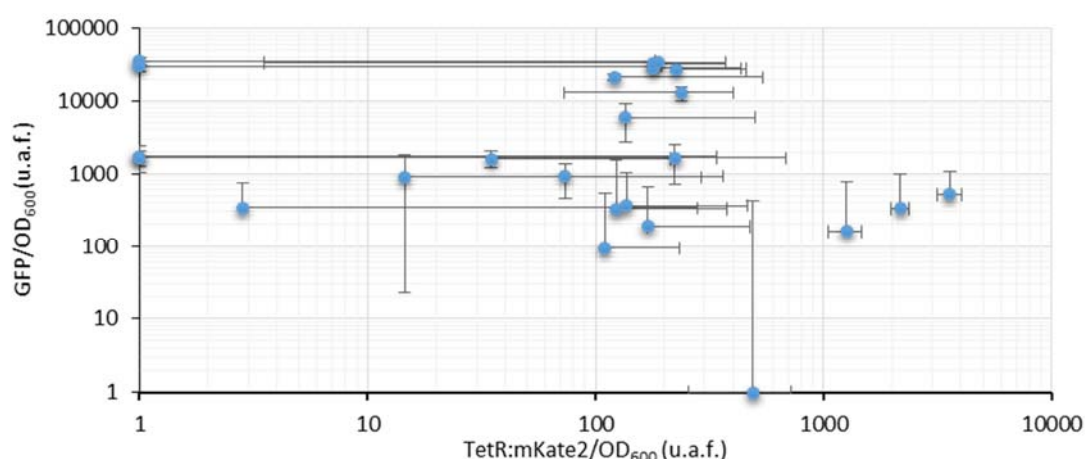


Figura 3-7. Actividad promotora de p_{tetA} (fluorescencia verde, eje de ordenadas) frente a concentración de represor TetR:mKate2 (fluorescencia roja, eje de abscisas) para la cepa con los vectores pRRG13 y 62. Los puntos azules representan la media de los valores en el estado estacionario, que se corresponden a los datos en la Figura 3-6, paneles A, B y C. Las desviaciones estándar asociadas a cada punto se representan en la dirección del eje correspondiente.

Los resultados mostrados hasta este punto revelan un problema de sensibilidad del fluorímetro. Uno de los objetivos marcados en esta tesis es la comparación de la eficiencia y precisión entre la fluorimetría de cultivo y la microscopía de epifluorescencia. Por consiguiente, los sistemas genéticos construidos posteriormente también se analizaron en fluorímetro y los resultados se detallan a continuación.

3.1.5. Dispositivo genético de medida de un solo vector e inserción cromosómica

Con la intención de anular el posible efecto producido por la fluctuación independiente en el número de copias de los vectores reportero y de expresión, se unificaron ambos sistemas en un solo vector. Además, se insertaron en el cromosoma de BW27783, para disminuir la fluctuación del número de copias total (Figura 3-8).

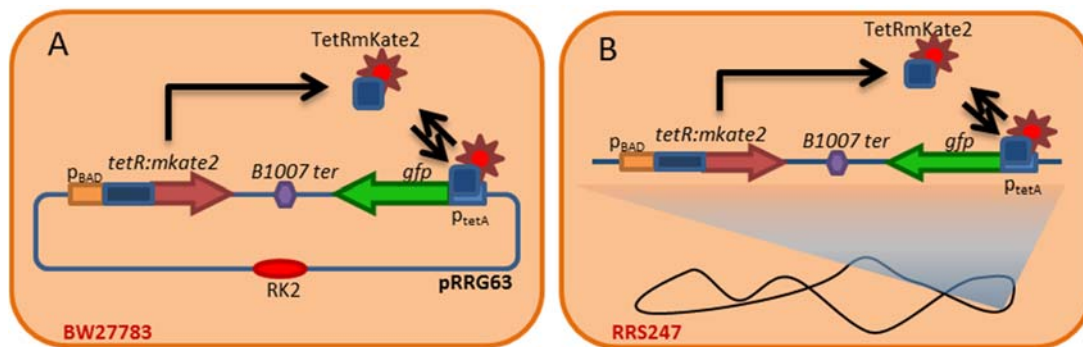


Figura 3-8. Esquema de las cepas experimentales. (A) BW27783 con el vector pRRG63. (B) Inserción en el cromosoma de BW27783 del *cassette* con los sistemas de expresión y reportero proveniente de pRRG63. La proteína de fusión TetR:mKate2 se produce a partir del promotor p_{BAD} y se une al promotor p_{tetA} , a partir del cual se transcribe el gen reportero *gfp*. Este *cassette* represor-reportero se encuentra codificado, bien en un vector con origen de replicación RK2 (óvalo rojo en panel A), o en el cromosoma bacteriano (panel B).

El diferente número de copias del sistema reportero en la bacteria repercute en el nivel de señal fluorescente producido por el mismo. Primero, se construyó como control un vector que llevaba tanto el sistema reportero como el de expresión, pero sin secuencia codificante de TetR (vector pRRG85). Este *cassette*, con la región reportera y de expresión, se incorporó al cromosoma de BW27783 (cepa RRS283), en el locus del operón de la arabinosa. Con estos dos nuevos constructos genéticos se midió la actividad de p_{tetA} en ausencia de represor (Figura 3-9) para poder comparar los distintos niveles de producción. Los resultados mostraron que la fluorescencia producida es proporcional al número de copias. pRRG85 mostró una producción de GFP mayor que pRRG13 (vector reportero de la actividad de p_{tetA}), y RRS283 una producción menor que ambos. Al comprobar la actividad en presencia de glucosa se observó un incremento en los tres sistemas frente a la fluorescencia registrada en el cultivo en glicerol, llegando a incrementarse hasta $\approx 30\%$ en pRRG85.

Por otro lado, los tiempos de generación se mantuvieron relativamente estables, manteniéndose entre 50 y 60 min en glicerol, y entre 45 y 55 min creciendo en glucosa.

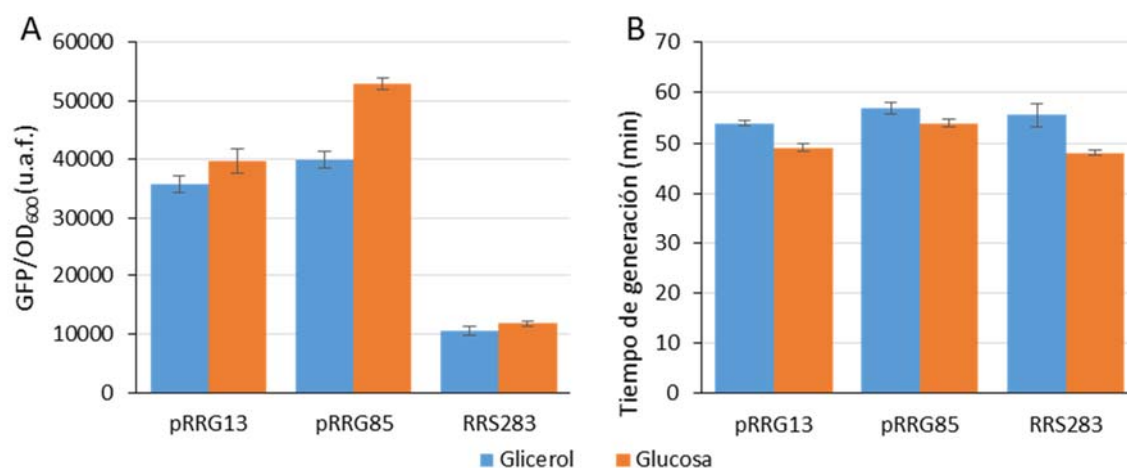


Figura 3-9. Cuantificación de la actividad de p_{tetA} en pRRG13, pRRG85 e insertado en el cromosoma (RRS283). Las cepas se cultivaron en M9 con glicerol (azul) o con glucosa (naranja) y se analizaron en el fluorímetro como se detalla en el apartado 2.7 de Materiales y Métodos (A) Valores medios de la actividad promotora (GFP/OD_{600}) durante el crecimiento exponencial del cultivo. (B) Tiempos de generación en min. Las barras de error reflejan la desviación estándar de las distintas medidas.

Conociendo los datos de actividad del promotor en ausencia de regulador, se analizaron a la caracterización de los mismos sistemas en presencia de TetR:mKate2. En primer lugar se construyó pRRG63, el vector reportero para la actividad de p_{tetA} y productor de TetR:mKate2.

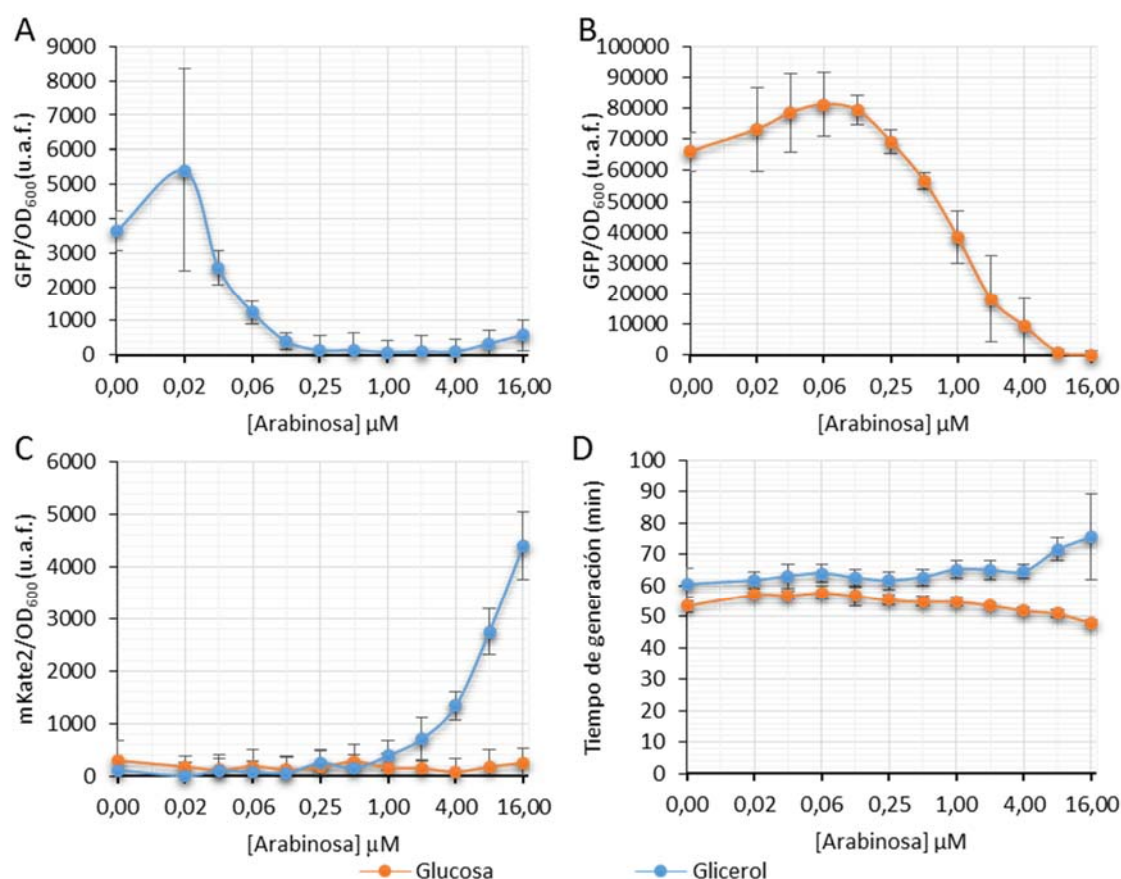


Figura 3-10. Fluorescencia roja y verde medida para BW27783 con pRRG63. Los cultivos crecieron en M9 con glicerol (azul) o glucosa (naranja) y se analizaron en el fluorímetro como se detalla en el apartado 2.7. Se representa la actividad promotora relativizada por el crecimiento (GFP/OD_{600}) frente al gradiente de arabinosa en cultivo: (A) en M9 con glicerol y (B) en M9 con glucosa. (C) Representación de la concentración de TetR:mKate2 producida relativizada por el crecimiento ($mKate2/OD_{600}$) frente al gradiente de arabinosa. (D) Tiempos de generación de los cultivos. Las barras de error representan la desviación estándar de los datos.

Como adelantaban los valores obtenidos con el promotor en ausencia de represor (Figura 3-9), las intensidades de fluorescencia verde obtenidas en todos los casos fueron mayores que las obtenidas con el sistema de los dos vectores (Figura 3-10 vs Figura 3-6). La fluorescencia roja, es decir, la cantidad de TetR:mKate2, se mantuvo en niveles similares en ambos sistemas (pRRG13+pRRG62 vs pRRG63), ya que el origen de replicación del vector de expresión pRRG62 era el mismo que el del vector dual pRRG63. Estos datos con intensidad mayor de fluorescencia verde permitieron, por lo tanto, observar un mayor rango de represión de la

actividad promotora, pero los valores menores continuaron por debajo del límite de detección del fluorímetro.

Representando en una gráfica, al igual que en la Figura 3-7, los datos de actividad de p_{tetA} medida en fluorescencia verde frente a la concentración de TetR medida en fluorescencia roja, se observó el mayor rango de regulación descrito. No obstante, se corroboró que la sensibilidad de la técnica no permitía correlacionar los niveles de regulador con la actividad promotora (Figura 3-11). Al igual que en el caso anterior, sólo cuando la señal verde se encontraba en valores de fondo, la señal roja superaba el umbral de detección.

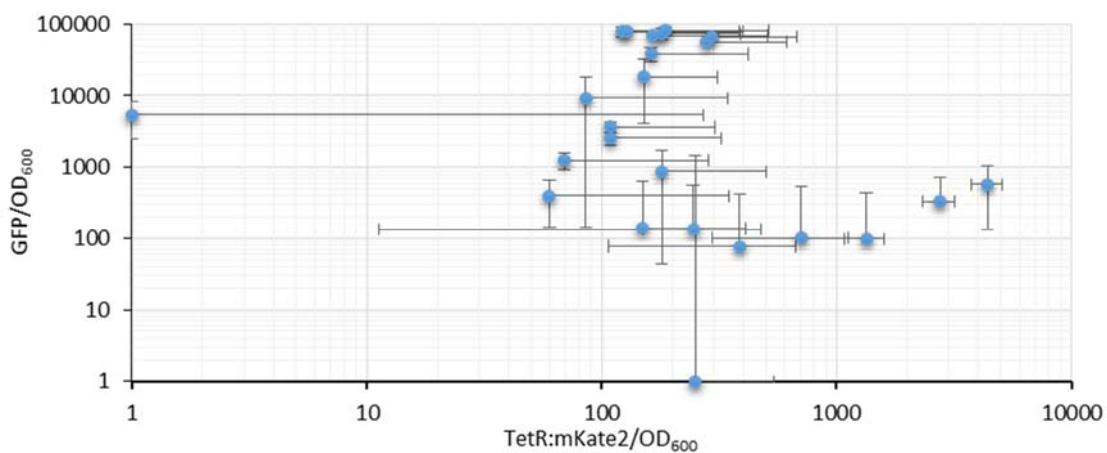


Figura 3-11. Actividad promotora de p_{tetA} (fluorescencia verde) frente a concentración de represor TetR:mKate2 (fluorescencia roja), ambos relativizados por el crecimiento, para la cepa con el vector pRRG63. Los puntos azules representan la media de los valores en el estado estacionario, que se corresponden a los datos en la Figura 3-10, paneles A, B y C. Las desviaciones estándar asociadas a cada punto se representan en la dirección del eje correspondiente.

El siguiente constructo que se analizó fue la inserción cromosómica de los sistemas reportero y de expresión procedentes de pRRG63, la cepa RRS247 (Figura 3-8.B). Se llevó a cabo el mismo procedimiento que con las construcciones anteriores, midiendo la fluorescencia en los canales rojo y verde de cultivos crecidos en M9 con glicerol y con glucosa en un gradiente de arabinosa. Los resultados obtenidos se muestran en la Figura 3-12. Observando el tiempo de generación, la cepa RRS247 tuvo un comportamiento similar a las demás, sin detectarse una deceleración en el crecimiento al inducir con arabinosa. Respecto a los datos fluorimétricos, el rango de actividad de p_{tetA} observado fue similar al visto con pRRG63, aunque mostrando una intensidad de la señal inferior, debida al menor número de copias. No obstante, la desviación estándar de la actividad promotora tuvo valores similares a lo largo del gradiente de arabinosa, es decir, se observó un comportamiento más homogéneo del sistema. Las medidas de la fluorescencia roja en esta ocasión no sobrepasaron el umbral de sensibilidad de la técnica, por

lo que los valores fluctuaron en torno a la señal de ruido en el canal, lo que impidió una vez más la correlación entre concentración de represor y nivel de actividad promotora.

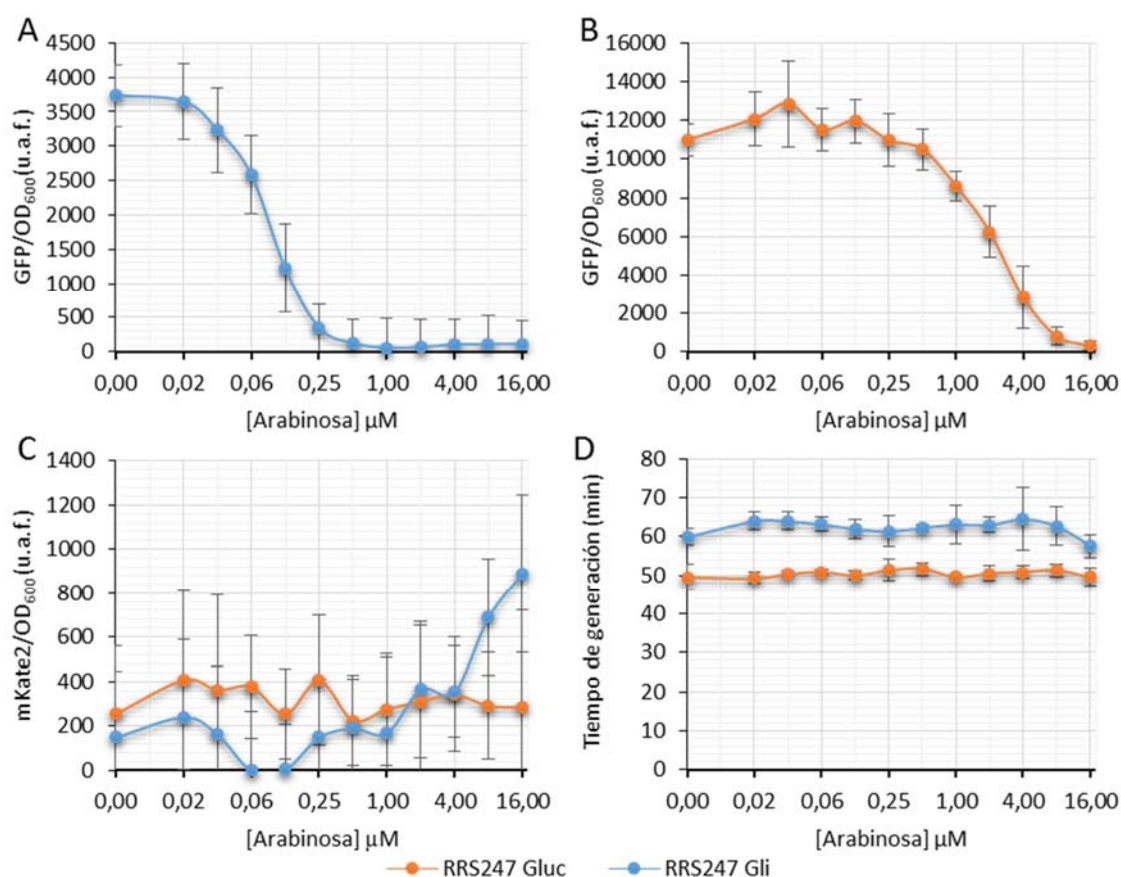


Figura 3-12. Fluorescencia roja y verde medida para la cepa RRS247. Los cultivos crecieron en M9 con glicerol (azul) o glucosa (naranja) y se analizaron en el fluorímetro como se detalla en el apartado 2.7. Se representa la actividad promotora relativizada por el crecimiento (GFP/OD_{600}) frente al gradiente de arabinosa en cultivo: (A) en M9 con glicerol y (B) en M9 con glucosa. (C) Representación de la concentración de TetR:mKate2 producida relativizada por el crecimiento ($mKate2/OD_{600}$) frente al gradiente de arabinosa. (D) Tiempos de generación de los cultivos. Las barras de error representan la desviación estándar de los datos.

3.1.6. Control de las proteínas fluorescentes: TetR:GFP y $p_{tetA}:mKate2$

Como comprobación tanto de la actividad de la proteína de fusión como de la sensibilidad de los canales de fluorescencia, se construyó una cepa idéntica a RRS247 pero intercambiando las proteínas fluorescentes. De esta manera la fluorescencia roja de mKate2 indicaría la actividad de p_{tetA} , mientras que la fluorescencia verde reflejaría la concentración de la proteína de fusión TetR:GFP (cepa RRS279).

Al igual que en los anteriores casos, se cultivó la cepa en el fluorímetro a 37 °C en M9 con glicerol y en presencia de un gradiente de concentraciones de arabinosa (Figura 3-13). La cepa no mostró efectos adversos de la inducción sobre el crecimiento. Los datos fluorimétricos

mostraron que, con esta conformación de las proteínas fluorescentes, la caracterización del sistema era más limitada en RRS279 respecto a RRS247. Dada la sensibilidad de los canales de fluorescencia, se observó una ligera mejoría en la definición de la concentración de TetR, pero con la contrapartida de reducir en gran medida la señal de la actividad del promotor p_{tetA} . No obstante, se pudo corroborar una inducción similar de TetR en función de la concentración de arabinosa y la represión de la actividad del promotor.

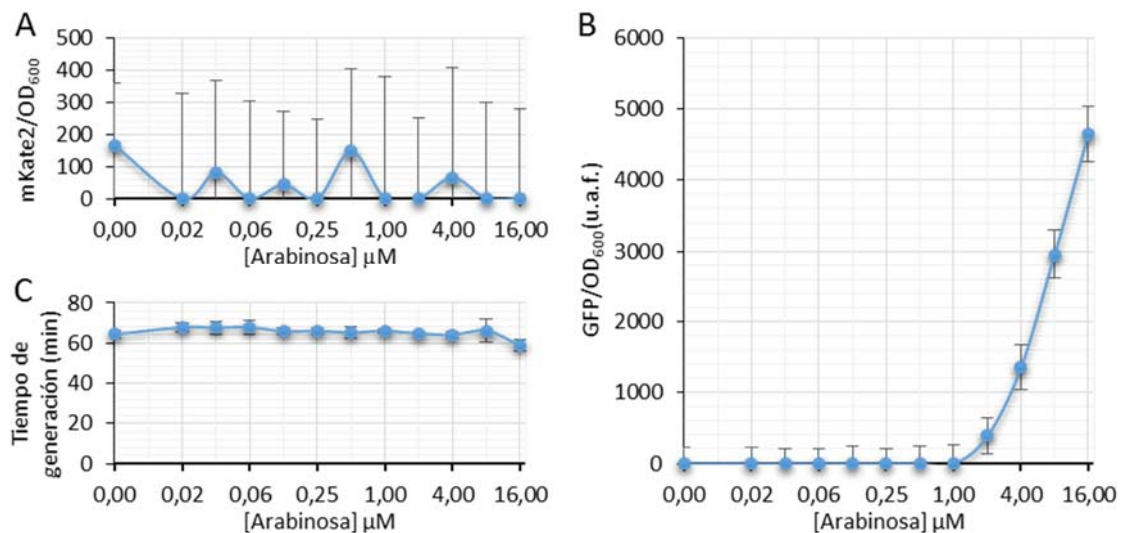


Figura 3-13. Fluorescencia roja y verde medida para la cepa RRS279. Los cultivos crecieron en M9 con glicerol y se analizaron en el fluorímetro como se detalla en el apartado 2.7. (A) Se representa la actividad promotora relativizada por el crecimiento ($mKate2/OD_{600}$) frente al gradiente de arabinosa. (B) Representación de la concentración de TetR:GFP producida relativizada por el crecimiento (GFP/OD_{600}) frente al gradiente de arabinosa. (C) Tiempos de generación de los cultivos. Las barras de error representan la desviación estándar de los datos.

3.2. Cuantificación de la regulación transcripcional mediante microscopía de epifluorescencia

Tras evidenciar las limitaciones de la fluorimetría de cultivo como método para cuantificar la regulación transcripcional, se optó por la microscopía de epifluorescencia en célula única como técnica de medida.

El propósito de esta nueva aproximación fue cuantificar un rango mayor tanto de la entrada como de la salida del sistema promotor/regulador, con el propósito de poder revelar su función de transferencia. Dado que, en la técnica de fluorimetría en cultivo, el mayor impedimento fue la cuantificación de la concentración de regulador, en este caso sólo se caracterizaron las fusiones traduccionales, para conseguir una medida directa de la misma. Siguiendo el protocolo explicado en el apartado 2.8 se midió la actividad represora de TetR sobre p_{tetA} mediante los vectores pRRG13 y pRRG62 en BW27783, el vector pRRG63 en BW27783 y de la cepa RRS247.

Previamente al análisis de los sistemas de medida, se midió la autofluorescencia de la cepa BW27783 en los canales de fluorescencia de las proteínas reporteras (Figura 3-14). Los valores de fluorescencia fueron, en el canal verde, un orden de magnitud mayores que en el rojo. BW27783, cultivada en M9 con glicerol, mostró una fluorescencia media de 149 ± 40 u.a.f. en el canal rojo y de 1642 ± 284 u.a.f. en el canal verde. Cultivada en glucosa, en cambio, la cepa mostró una fluorescencia media de 232 ± 96 u.a.f. en el canal rojo y de 1204 ± 203 u.a.f. en el canal verde.

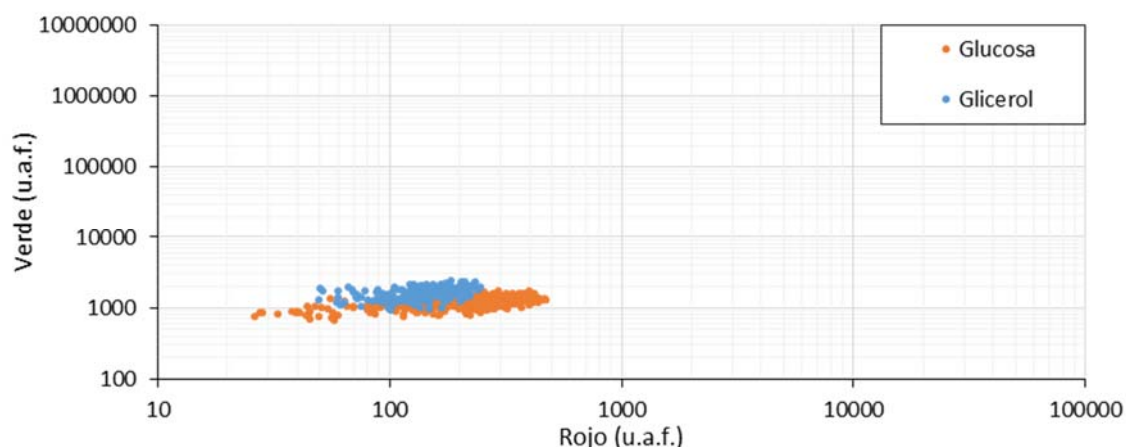


Figura 3-14. Autofluorescencia de la cepa BW27783 en el microscopio de epifluorescencia. Cada punto representa los valores en el canal verde y rojo de una célula individual. El cultivo de la cepa y la toma de imágenes se realizaron como se indica en el apartado 2.8.1. Se analizó tanto el cultivo en M9 con glicerol (azul) como en M9 con glucosa (naranja).

Los datos obtenidos para las poblaciones de BW27783 se utilizaron como referencia para las cuantificaciones posteriores de los sistemas de medida de regulación transcripcional, pero no se restaron a los valores conseguidos en el tratamiento de datos.

3.2.1. Sistema de dos vectores: pRRG13 y pRRG62 en BW27783

El principal beneficio esperado al cambiar de técnica era la mejor cuantificación de la concentración de molécula reguladora. En este caso, como no se pretendía inferir su concentración sino cuantificarla directamente, se optó por una serie de combinaciones de M9 con glicerol o glucosa y con distintas concentraciones de arabinosa para inducir *tetR::mkate2*. De esta manera se pretendió obtener el mayor rango continuo de concentraciones de regulador en el cual p_{tetA} mostrase respuesta, es decir, a efectos experimentales, inducir un rango de intensidades de fluorescencia roja con sus estados intermedios completamente muestreados. Siguiendo el protocolo experimental previamente descrito, los cultivos fueron crecidos en M9 líquido con los antibióticos y las condiciones de inducción deseadas, para luego ser sembrados en *pads* de agarosa. Se realizaron micrografías de dichos *pads* en contraste de

fases y fluorescencia roja y verde. Tras el análisis de las imágenes se obtuvieron los resultados detallados a continuación.

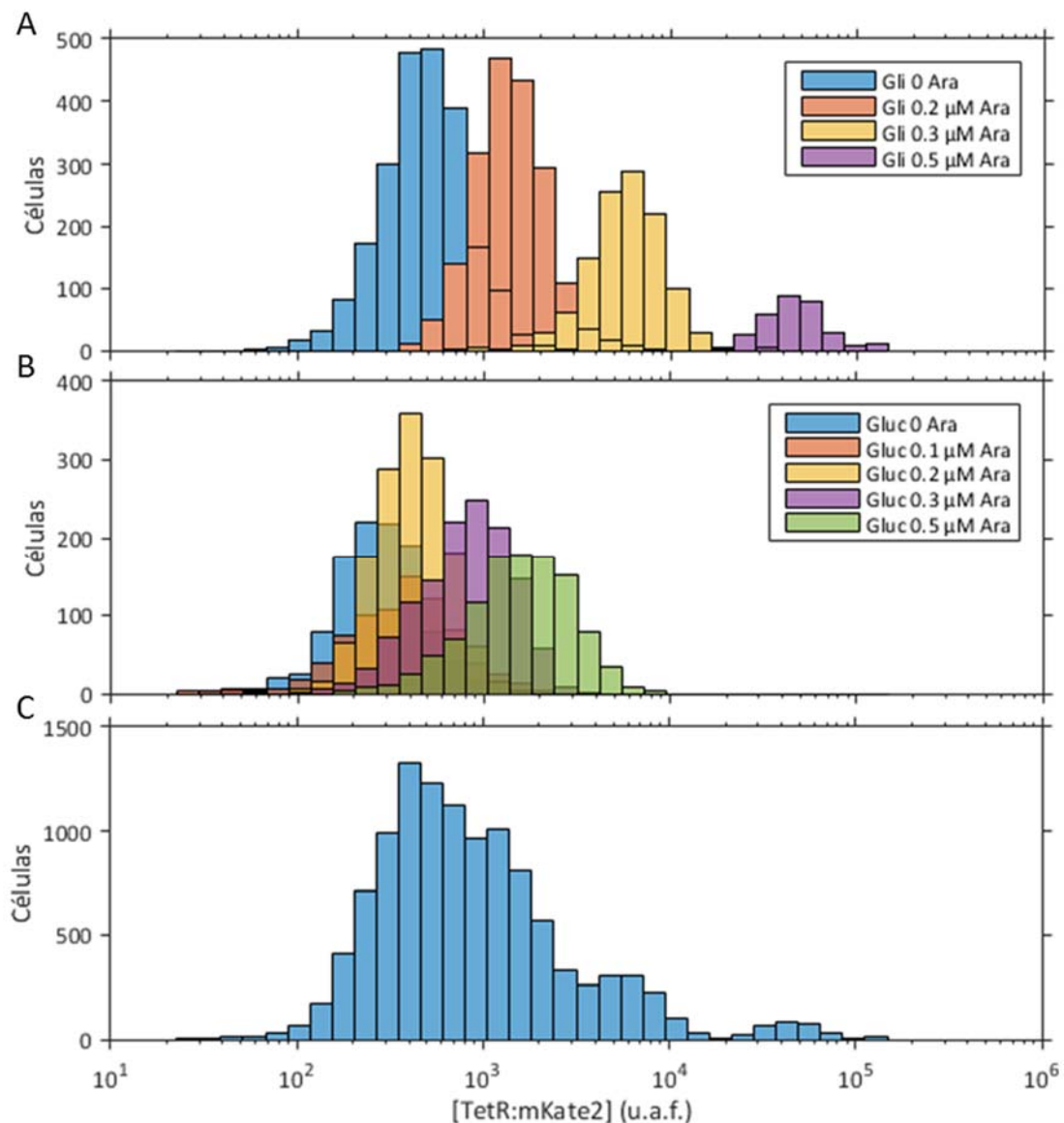


Figura 3-15. Histograma del número de células analizadas con el sistema pRRG13+pRRG62, en función de su fluorescencia roja (concentración de TetR:mKate2). La medición se realizó como se indica en el apartado 2.8.1 de Materiales y Métodos. (A) Representación de cada población obtenida mediante cultivo en M9 con glicerol y diferentes concentraciones de arabinosa. (B) Representación de las poblaciones cultivadas en M9 con glucosa y diferentes concentraciones de arabinosa. (C) Suma de todas las células. Las diferentes condiciones de inducción se detallan en las leyendas de las gráficas con un código de color.

Tal y como se muestra en la Figura 3-15, se realizaron distintas inducciones del promotor p_{BAD} consiguiendo diferentes concentraciones de TetR:mKate2, medidas experimentalmente mediante fluorescencia roja. Las mayores intensidades se consiguieron en M9 con glicerol y arabinosa 0,5 μ M. La población con la distribución menos intensa en el rojo se consiguió mediante cultivo en M9 con glucosa sin arabinosa. No obstante, las poblaciones inducidas en M9 con glucosa y arabinosa 0,1 y 0,2 μ M, y la población cultivada en glicerol sin arabinosa,

obtuvieron distribuciones prácticamente solapantes, cercanas a la crecida en M9 con glucosa sin arabinosa.

Valores de fluorescencia verde fueron registrados al mismo tiempo que las imágenes de fluorescencia roja para esas mismas células. Representando dicha fluorescencia verde (actividad promotora) frente a la fluorescencia roja (concentración de TetR) se pudieron correlacionar los dos componentes del sistema regulador (Figura 3-16).

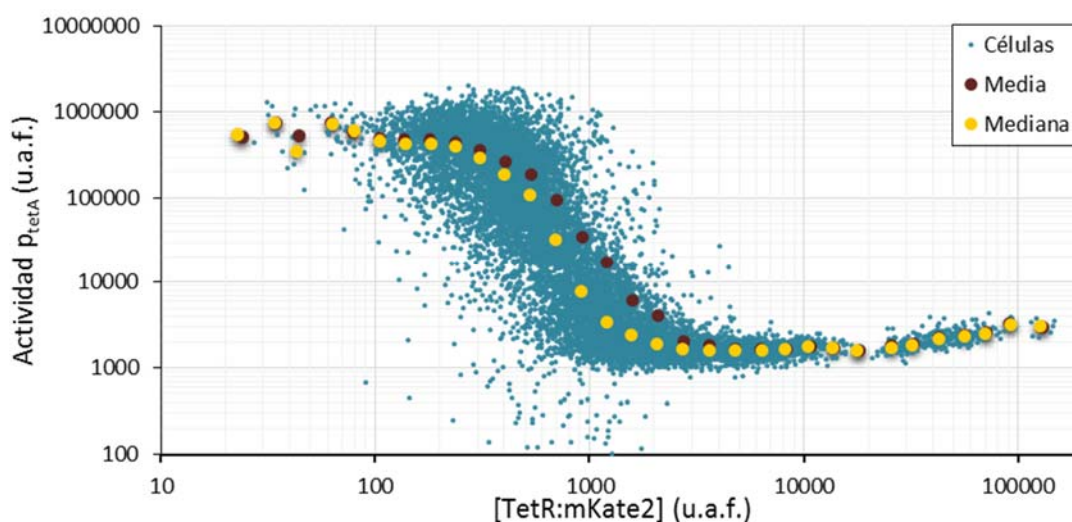


Figura 3-16. Actividad promotora de p_{tetA} (fluorescencia verde) en función de la concentración de TetR:mKate2 (fluorescencia roja) en la cepa con los vectores pRRG13 y pRRG62. El cultivo de la cepa y la toma de imágenes se realizaron como se indica en el apartado 2.8.1 de Materiales y Métodos. Los niveles de inducción utilizados se describen en la Figura 3-15. Se representan la media (en puntos rojos) y la mediana (en puntos amarillos) correspondientes a cada uno de los *bins* en los que se dividió la señal en el canal rojo. Cada punto azul representa una célula

Analizando la correlación se observó que el incremento de la concentración de TetR produjo una represión en la actividad promotora de p_{tetA} . Las bacterias sin inducción del represor presentaron la mayor actividad transcripcional. El incremento de represor (de 100 a 1000 u.a.f.) desplazó la fluorescencia de las bacterias a una zona de incertidumbre entre valores intermedios de actividad. Finalmente, las concentraciones mayores de TetR reprimieron más el promotor, alcanzando el estado mínimo de actividad, donde las bacterias se comportaron de una manera más uniforme. Se calcularon la media y la mediana de los valores de fluorescencia a intervalos logarítmicos de concentración de regulador para describir de una manera sencilla el comportamiento caracterizado. Dada la dispersión de actividad promotora que presentaron las diferentes células para un mismo nivel de regulador, la mediana se reveló como un descriptor más adecuado.

Estos datos permitieron estimar la función de transferencia del mecanismo de regulación transcripcional para el sistema constituido con los vectores pRRG13 y pRRG62. Apoyándose en

el marco teórico descrito en (Fernandez-Lopez *et al.*, 2010), se calcularon los parámetros descriptivos de la función reguladora (Ecuación 3-1):

$$\frac{Y_{max}}{Y_{SS}} = \frac{K^n + X_{SS}^n}{K^n + (X_{SS}^n / R_Y^X)} \quad \text{donde} \rightarrow R_Y^X = \frac{Y_{max}(X)}{Y_{min}(X)}$$

Ecuación 3-1. Función de transferencia para un represor transcripcional.

Según este modelo, la dinámica de producción del promotor diana (Y) responde a la concentración de regulador (X). En el estado estacionario (SS), el grado de represión (Y_{max}/Y_{SS}) es función de la concentración de regulador (X_{SS}). K representa la constante de disociación y n es el grado de cooperatividad aparente en la dinámica observada. R se define como el rango regulatorio del promotor dado el represor.

Los valores experimentales se ajustaron a la Ecuación 3-1 mediante el algoritmo Levenberg-Marquardt para regresiones no lineales. Como se observó previamente, la media y la mediana de las nubes de puntos obtenidas difirieron en la descripción de la dinámica del promotor por lo que se ajustó la ecuación de transferencia con ambos descriptores.

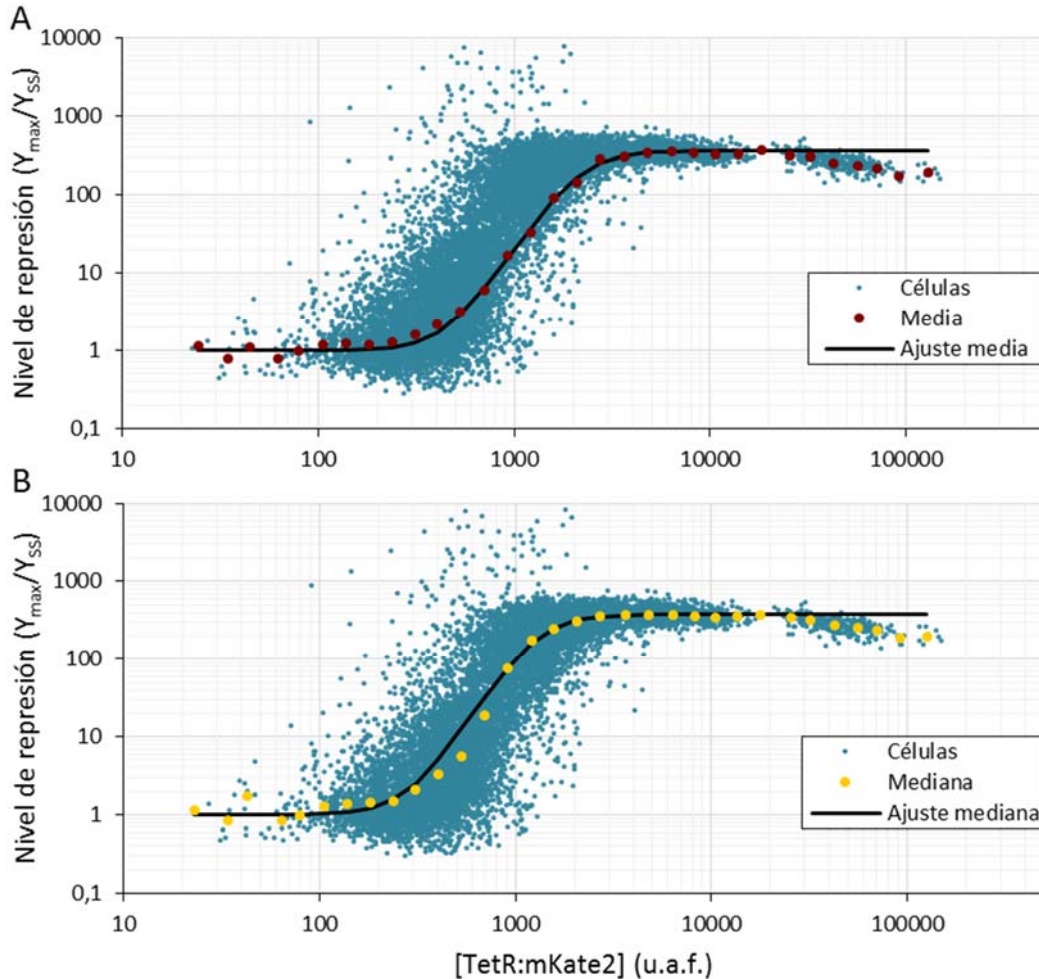


Figura 3-17. Nivel de represión de p_{tetA} en función de la concentración de TetR para la cepa con los vectores pRRG13 y pRRG62. Se representa el nivel de represión de la actividad de p_{tetA} (Y_{max}/Y_{ss}) frente a los valores de fluorescencia roja de TetR:mKate2, partiendo de los datos de la población de la Figura 3-16. Los valores de represión se ajustaron a la Ecuación 3-1. (A) Ajuste de la media (en rojo) de los datos. (B) Ajuste de la mediana (en amarillo) de los datos. La función obtenida mediante el ajuste se representa como una línea negra. Se muestran los valores de las células individuales como muestra de la dispersión del comportamiento del sistema (puntos azules).

Para el caso de las medias de los valores la regresión dio una K de 441,69 u.a.f. El rango regulatorio R fue 361,8 y el índice de cooperatividad n fue 3,686. El coeficiente R^2 de la curva ajustada fue 0,9932. Asimismo, para las medianas de los valores, la K fue 276,94 u.a.f., el rango regulatorio R fue 379,2 y n fue 3,828, mientras que el ajuste tuvo una R^2 de 0,9947. Las curvas predichas mediante estos valores se muestran en la Figura 3-17.

3.2.2. Sistema de un vector: pRRG63 en BW27783

Al igual que en caso anterior, siguiendo el protocolo experimental descrito en el apartado 2.8.1 los cultivos fueron crecidos en M9 líquido con los antibióticos y las condiciones de inducción deseadas. Se sembraron en *pads* y se les realizaron fotografías en contraste de fases y fluorescencia roja y verde.

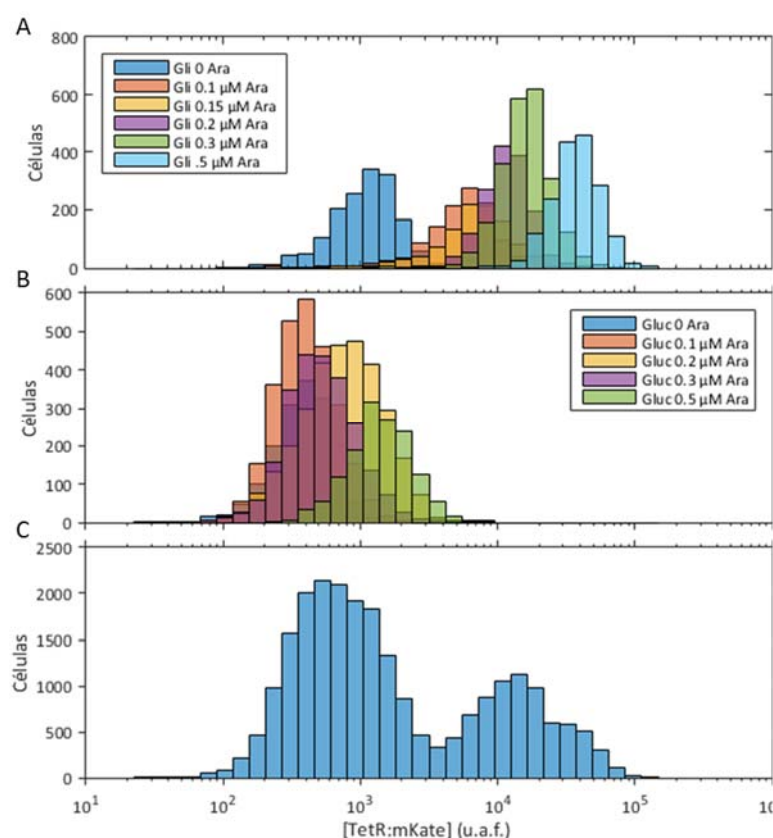


Figura 3-18. Histograma del número de células analizadas con el vector pRRG63 en función de su fluorescencia roja (concentración de TetR:mKate2). La medición se realizó como se indica en el apartado 2.8.1 de Materiales y Métodos. (A) Representación de cada población obtenida mediante cultivo en M9 con glicerol y diferentes concentraciones de arabinosa. (B) Representación de las poblaciones cultivadas en M9 con glucosa y diferentes concentraciones de arabinosa. (C) Suma de todas las células. Las diferentes condiciones de inducción se detallan en las leyendas de las gráficas con un código de color.

La Figura 3-18 muestra el rango de concentraciones de regulador TetR:mKate2 conseguido mediante las distintas condiciones inductoras. Como en el caso anterior, las bacterias se desplazaron de menor a mayor fluorescencia roja a medida que el sistema inductor produjo

TetR. El nivel de concentraciones obtenido fue muy similar al producido por el vector pRRG62 ya que ambos vectores comparten origen de replicación. Asimismo, se comprobó que las regiones intermedias del rango estuvieran muestreadas.

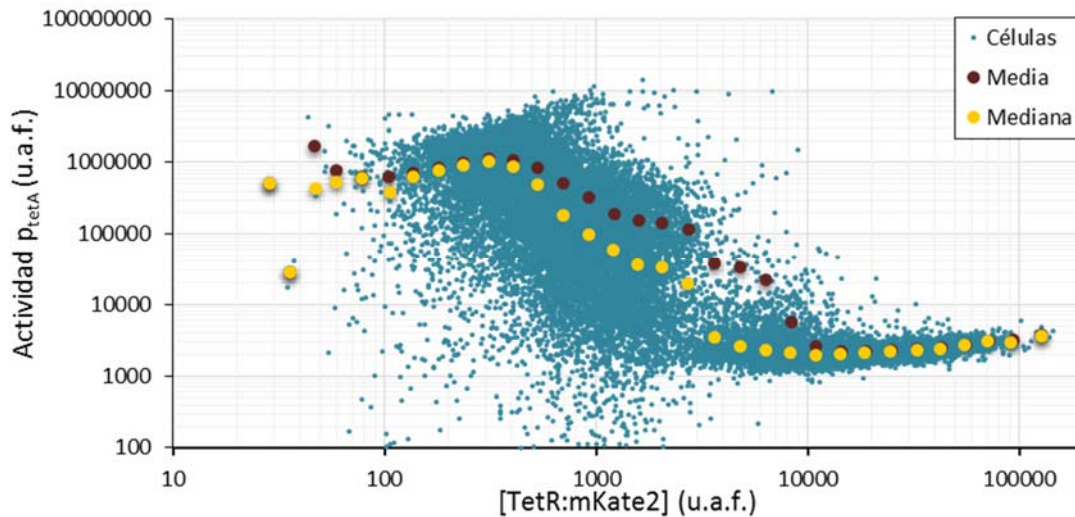


Figura 3-19. Actividad promotora de p_{tetA} (fluorescencia verde) en función de la concentración de TetR:mKate2 (fluorescencia roja) en la cepa con el vector pRRG63. El cultivo de la cepa y la toma de imágenes se realizaron como se indica en el apartado 2.8.1 de Materiales y Métodos. Los niveles de inducción utilizados se describen en la Figura 3-18. Se representan la media (en puntos rojos) y la mediana (en puntos amarillos) correspondientes a cada uno de los bins en los que se dividió la señal en el canal rojo. Cada punto azul representa una célula.

La represión de TetR sobre la actividad transcripcional de p_{tetA} quedó de nuevo patente. No obstante, se percibió un rango más amplio de actividad, con los niveles superiores de en torno a un orden de magnitud mayores. Esto es debido, como ya vimos con los datos fluorimétricos en cultivo, por el incremento del número de copias de p_{tetA} en pRRG63 respecto a pRRG13. Igual que con el par de vectores pRRG13/62 la media presentó bastante desplazamiento frente a la mediana, principalmente en las concentraciones intermedias de regulador por la gran heterogeneidad en la actividad promotora registrada para esas condiciones. Además se observó un posible estado intermedio de actividad promotora, cercano a la represión total, reflejado por un codo en la zona descendente de la curva. Los datos de célula única sugieren que, en dicho rango de concentraciones de TetR, el promotor escapa de la represión completa, generando un bajo pero detectable nivel de expresión, con producción de GFP, lo que hizo que los descriptores estadísticos reflejaran ese codo.

La media y la mediana de los valores se ajustaron a la Ecuación 3-1. La regresión de la media dio una K de 1324,6 u.a.f. El rango regulatorio R fue 498,3 y el índice de cooperatividad n fue 3,325. El coeficiente R^2 de la curva ajustada fue 0,9682. La regresión para la mediana, por su parte, dio una K de 651,67 u.a.f., el rango regulatorio R fue 537,7 y n fue 3,591, mientras que

el ajuste tuvo una R^2 de 0,9673. Las curvas predichas por la función de transferencia con los parámetros conseguidos para la media y la mediana se muestran en la Figura 3-20.

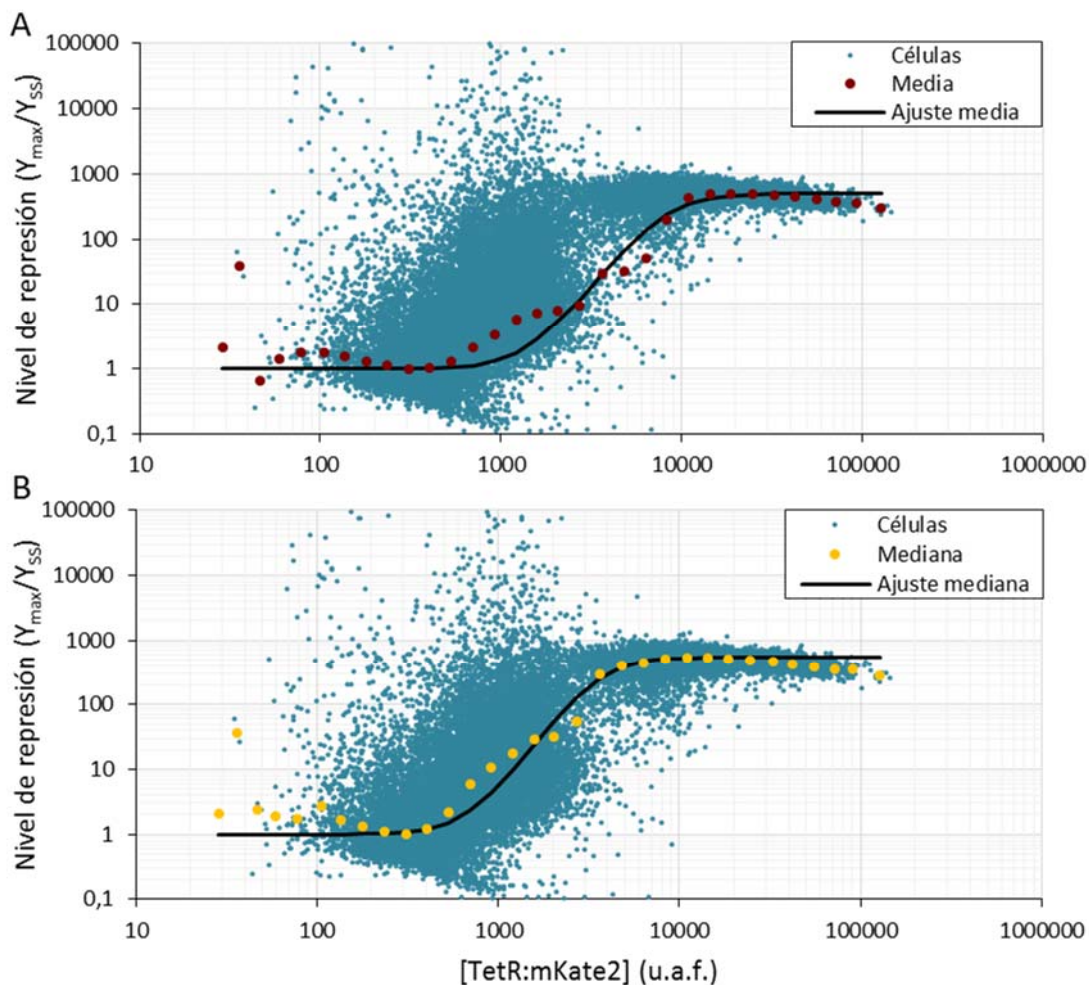


Figura 3-20. Nivel de represión de p_{tetA} en función de la concentración de TetR con el vector pRRG63. Se representa el nivel de represión de la actividad de p_{tetA} (Y_{max}/Y_{SS}) frente a los valores de fluorescencia roja de TetR:mKate2, partiendo de los datos de la población de la Figura 3-19. Los valores de represión se ajustaron a la Ecuación 3-1. (A) Ajuste de la media (en rojo) de los datos. (B) Ajuste de la mediana (en amarillo) de los datos. La función obtenida mediante el ajuste se representa como una línea negra. Se muestran los valores de las células individuales como muestra de la dispersión del comportamiento del sistema (puntos azules).

Teniendo presente la forma bifásica de la curva, se decidió también realizar un análisis como si se tratara de dos eventos de regulación para comprobar que el mecanismo regulatorio se comportaba de una manera equivalente a los previos y su cooperatividad aparente era similar. Se separaron los datos para cada posible evento regulatorio y se ajustaron a la función de transferencia independientemente (Figura 3-21). De esta manera, las medias del primer tramo de la curva, desde la máxima actividad promotora hasta el estado intermedio (Figura 3-21.A), obtuvieron una K de 650,14 u.a.f., R fue 7,8 y el índice de cooperatividad n fue 4,546. El coeficiente R^2 de la curva ajustada fue 0,9972. El segundo tramo, desde el estado intermedio

hasta la máxima represión (Figura 3-21.B), se ajustó a una K de 2851,73 u.a.f., R fue 63,35 y el índice de cooperatividad n fue 3,7, mientras que el coeficiente R^2 del ajuste fue 0,9348.

Las medianas de los datos (Figura 3-21.C y D), por su parte, dieron otros valores para los parámetros. El primer tramo de la curva dio una K de 501,21 u.a.f., su R fue 31,56 y n fue 4,508. El coeficiente R^2 del tramo ajustado fue 0.9853. Asimismo, el segundo tramo se ajustó a una K de 1830,54 u.a.f., el rango R fue 18,3 y el índice n fue 3,858, mientras que el coeficiente R^2 del ajuste fue 0,9666.

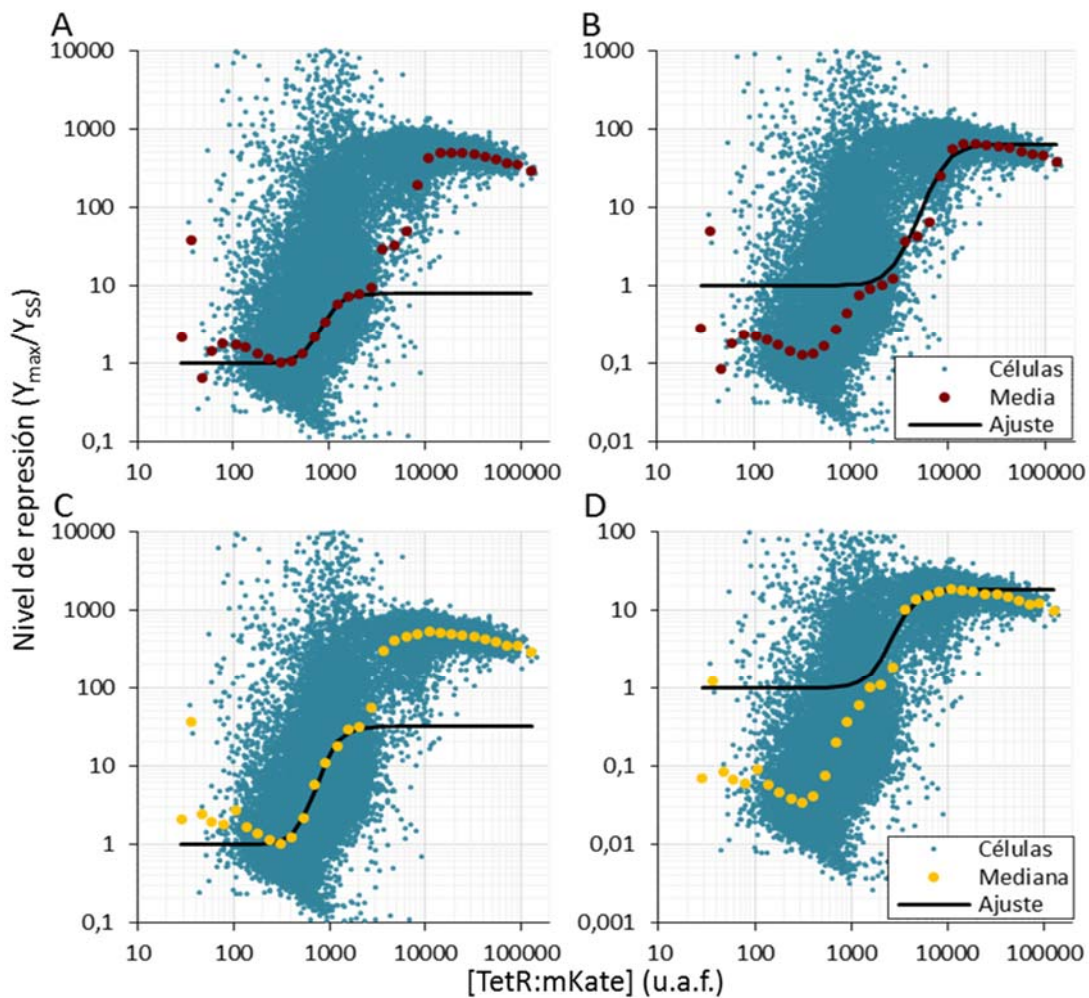


Figura 3-21. Nivel de represión de p_{tetA} en función de la concentración de TetR con el vector pRRG63. Los valores de la Figura 3-21 se ajustaron a la Ecuación 3-1 para dos posibles eventos de regulación. Se representa el nivel de represión de la actividad de p_{tetA} (Y_{max}/Y_{SS}) frente a los valores de regulador TetR:mKate2. (A) Ajuste de la media al primer tramo de la curva de regulación. (B) Ajuste de la media al segundo tramo de la curva. (C) Ajuste de la mediana al primer tramo de la curva de regulación. (D) Ajuste de la mediana al segundo tramo de la curva. La media se dibuja en rojo mientras que la mediana en amarillo. Cada punto azul representa una célula. La función ajustada se muestra como una línea negra.

3.2.3. Inserción cromosómica del sistema: cepa RRS247

La caracterización de la cepa RRS247 se realizó siguiendo el mismo procedimiento de los casos anteriores. Se realizaron cultivos en distintas condiciones inductoras y se sembraron en *pads* de agarosa para su medida en el microscopio. Asimismo, se realizaron micrografías en contraste de fases y fluorescencia roja y verde.

Una vez más, se analizó el rango de concentraciones de regulador conseguido mediante las distintas inducciones (Figura 3-22). En esta cepa el mecanismo de expresión de TetR tiene el menor número de copias por su localización cromosómica (de 1 a 4). Este menor número de copias produjo que las concentraciones máximas de TetR quedaran alrededor de un orden de magnitud por debajo de los casos precedentes.

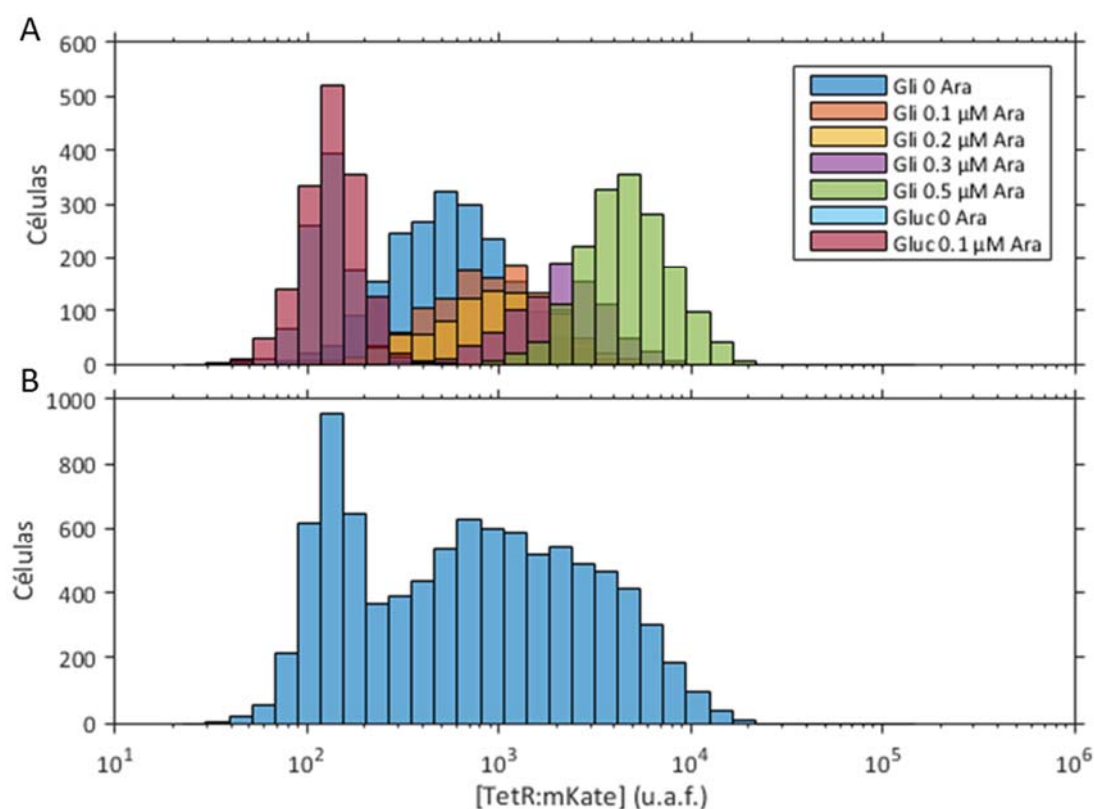


Figura 3-22. Histograma del número de células analizadas de la cepa RRS247 en función de su fluorescencia roja (concentración de TetR:mKate2). La medición se realizó como se indica en el apartado 2.8.1 de Materiales y Métodos. (A) Representación de cada población obtenida mediante cultivo en diferentes condiciones de inducción y (B) la suma de todas las células.

La correlación entre la fluorescencia roja y verde indicó, también en esta cepa, la represión transcripcional de TetR sobre la actividad de p_{tetA} . Al igual que con los niveles de expresión de TetR, la actividad máxima mostrada por el promotor en esta cepa quedó por debajo de los casos anteriores, y por lo tanto el rango descrito fue también menor. La dinámica de la represión en este caso indicó, más claramente que en pRRG63, un estado intermedio de

actividad de p_{tetA} en concentraciones de TetR cercanas a la represión total. Dada la heterogeneidad en la actividad para un mismo nivel de regulador, ese estado intermedio resultó más visible en el tratamiento de la media, mientras que al analizar la mediana su valor se mostró cercano al mínimo de actividad detectable.

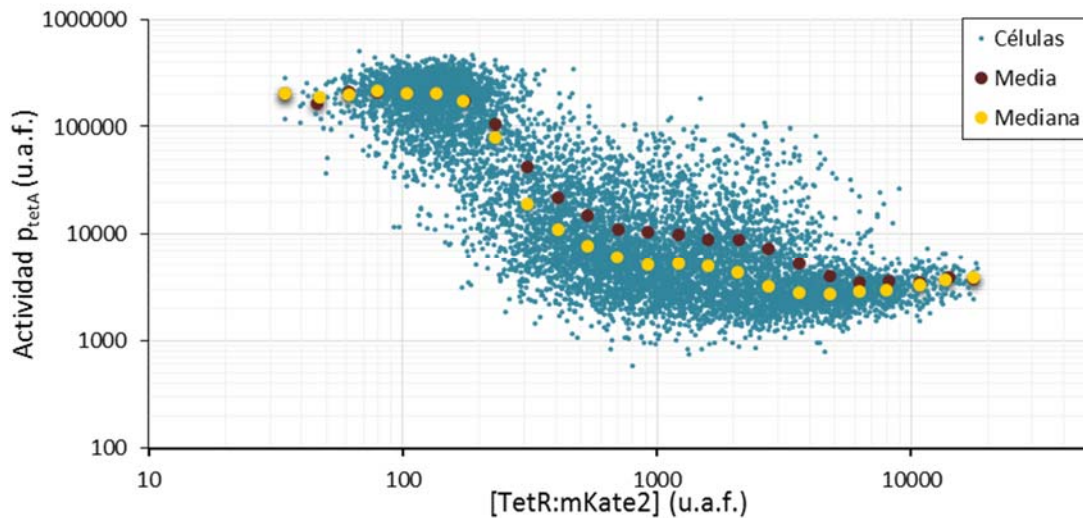


Figura 3-23. Actividad promotora de p_{tetA} (fluorescencia verde) en función de la concentración de TetR:mKate2 (fluorescencia roja) en la cepa RRS247. El cultivo de la cepa y la toma de imágenes se realizaron como se indica en el apartado 2.8.1 de Materiales y Métodos. Los niveles de inducción utilizados se describen en la Figura 3-22. Se representan la media (en puntos rojos) y la mediana (en puntos amarillos) correspondientes a cada uno de los bins en los que se dividió la señal en el canal rojo. Cada punto azul representa una célula.

La heterogénea actividad y las consiguientes medias y medianas representaron una aparente doble sigmoidea, que el modelo de la Ecuación 3-1 no era capaz de describir. No obstante, al igual que con el sistema del vector pRRG63, la curva trazada por los estadísticos se ajustó a la función de transferencia tratándola como un solo mecanismo de represión y como dos mecanismos independientes.

En primer lugar se ajustó la curva completa descrita tanto por la media como por la mediana de los valores (Figura 3-24). La regresión calculada a partir de la media dio una K de 113,03 u.a.f. El rango regulatorio R fue 60,7 y el índice de cooperatividad n fue 1,441. El ajuste tuvo un coeficiente R^2 de 0,9511. La regresión para la mediana, por su parte, devolvió una K de 54,57 u.a.f., el rango regulatorio R fue 80,2 y el índice n fue 1,486. El ajuste de la curva tuvo una R^2 de 0,9702.

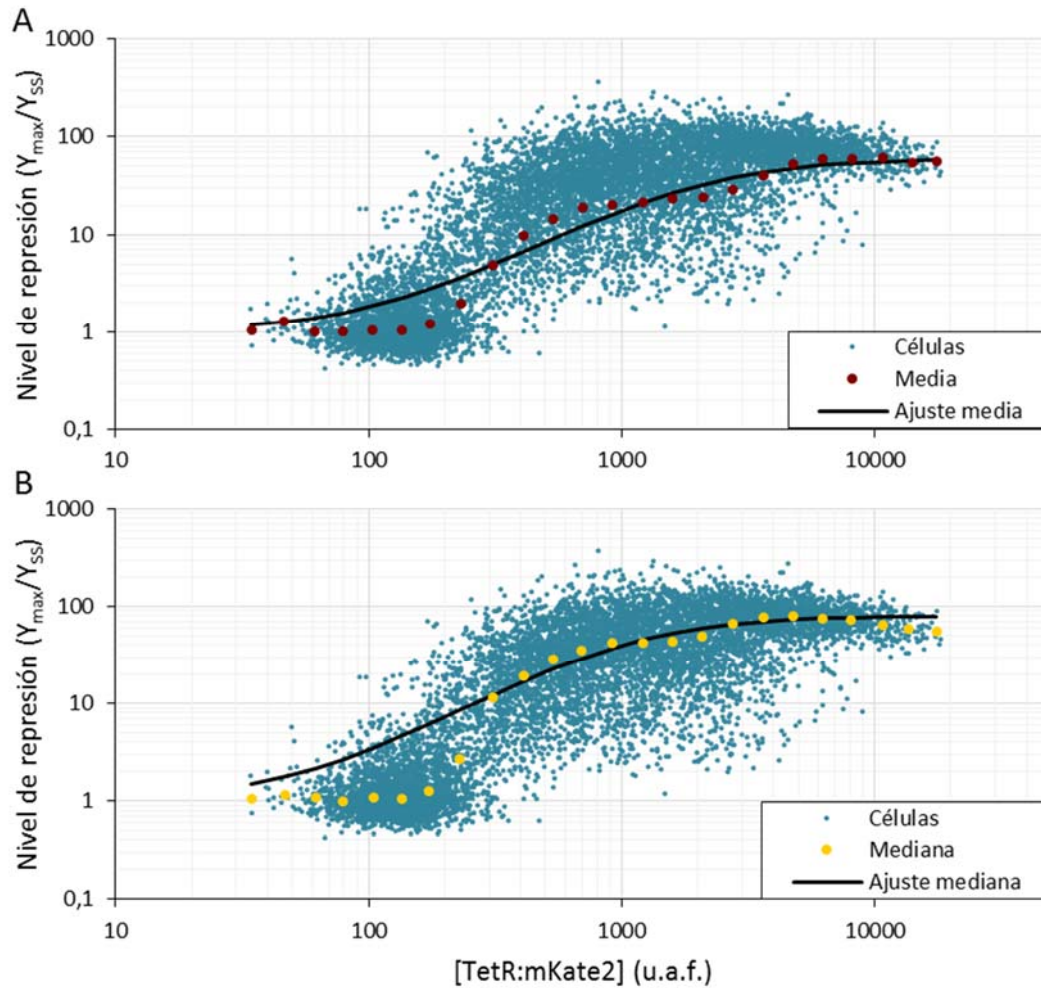


Figura 3-24. Nivel de represión de p_{tetA} en función de la concentración de TetR en la cepa RRS247. Se representa el nivel de represión de la actividad de p_{tetA} (Y_{max}/Y_{SS}) frente a los valores de fluorescencia roja de TetR:mKate2, partiendo de los datos de la población de la Figura 3-23. Los valores de represión se ajustaron a la Ecuación 3-1. (A) Ajuste de la media (en rojo) de los datos. (B) Ajuste de la mediana (en amarillo) de los datos. La función obtenida mediante el ajuste se representa como una línea negra. Se muestran los valores de las células individuales como muestra de la dispersión del comportamiento del sistema (puntos azules).

A continuación se analizó la dinámica como si se tratase de dos procesos de regulación, representado en la Figura 3-25. Las medias y medianas para cada posible evento regulatorio se ajustaron a la función de transferencia independientemente. Las medias del primer tramo de la curva (Figura 3-25.A) obtuvieron una K de 185,08 u.a.f., una R de 23,8 y una n de 3,262. El coeficiente R^2 de la curva ajustada fue 0,9953. El segundo tramo (Figura 3-25.B) se ajustó a una K de 2690,3 u.a.f., una R de 2,86 y una n de 3,655, mientras que el coeficiente R^2 del ajuste fue 0,9631.

La curva de las medianas se ajustó a la función de transferencia mediante los siguientes parámetros. El primer tramo de la curva (Figura 3-25.C) dio una K de 159,27 u.a.f., una R de 43,4 y una n de 3,648, con un coeficiente de determinación R^2 de 0,997. Del mismo modo, el

segundo tramo (Figura 3-25.D) se ajustó a una K de 2131,07 u.a.f., una R de 1,93 y un índice n de 3,858, con una R^2 de 0,9259.

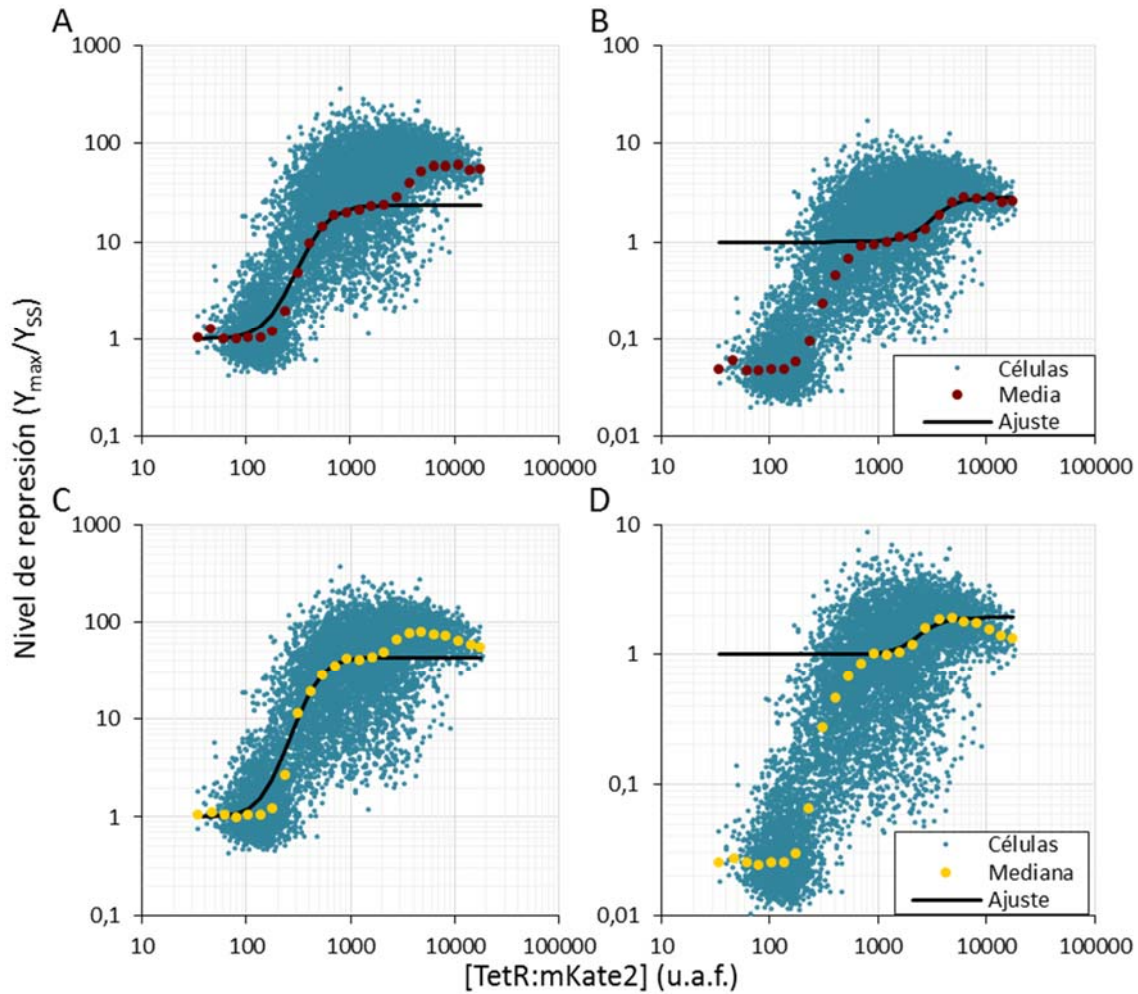


Figura 3-25. Nivel de represión de p_{tetA} en función de la concentración de TetR en la cepa RRS247. Los valores de la Figura 3-24 se ajustaron a la Ecuación 3-1 para dos posibles eventos de regulación. Se representa el nivel de represión de la actividad de p_{tetA} (Y_{max}/Y_{SS}) frente a los valores de regulador TetR:mKate2. (A) Ajuste de la media al primer tramo de la curva de regulación. (B) Ajuste de la media al segundo tramo de la curva. (C) Ajuste de la mediana al primer tramo de la curva de regulación. (D) Ajuste de la mediana al segundo tramo de la curva. La media se dibuja en rojo mientras que la mediana en amarillo. Cada punto azul representa una célula. La función ajustada se muestra como una línea negra.

Recapitulando, gracias a estos ensayos se corroboró que el sistema de regulación p_{tetA} /TetR se comporta como un mecanismo de represión. La metodología permitió probar una disminución de hasta 3 órdenes de magnitud de la actividad promotora en función de la concentración del represor. No obstante, la actividad transcripcional para cada concentración de regulador fue heterogénea (Figura 3-26), es decir, se dieron concentraciones de TetR que produjeron rangos amplios de concentración de GFP. Consecuentemente, las diferentes aproximaciones teóricas para la descripción del mecanismo, y sus posteriores predicciones del comportamiento, sólo permitieron una determinación tosca del estado de la población. Trasladando esta situación al

nivel celular, la bacteria no es capaz de definir un nivel concreto de actividad promotora frente a una concentración definida de regulador, pero tiene que ser capaz de sentir la señal y dar la respuesta adecuada. En el caso que nos ocupa, el operón *tet* debe detectar la concentración de tetraciclina y dar una adecuada respuesta produciendo la proteína de resistencia antibiótica. Por ello, procedimos al estudio de la transferencia de la información a través del mecanismo regulatorio desde el punto de vista de la teoría de la información.

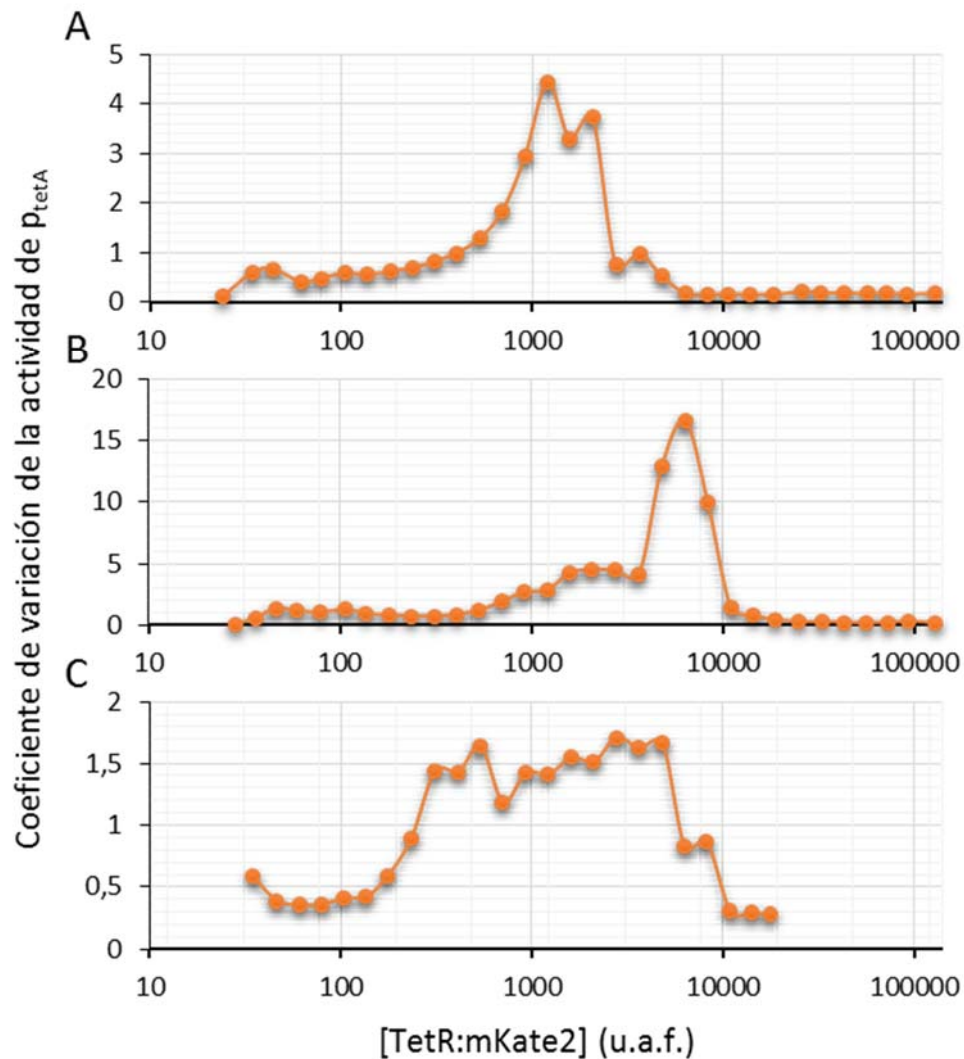


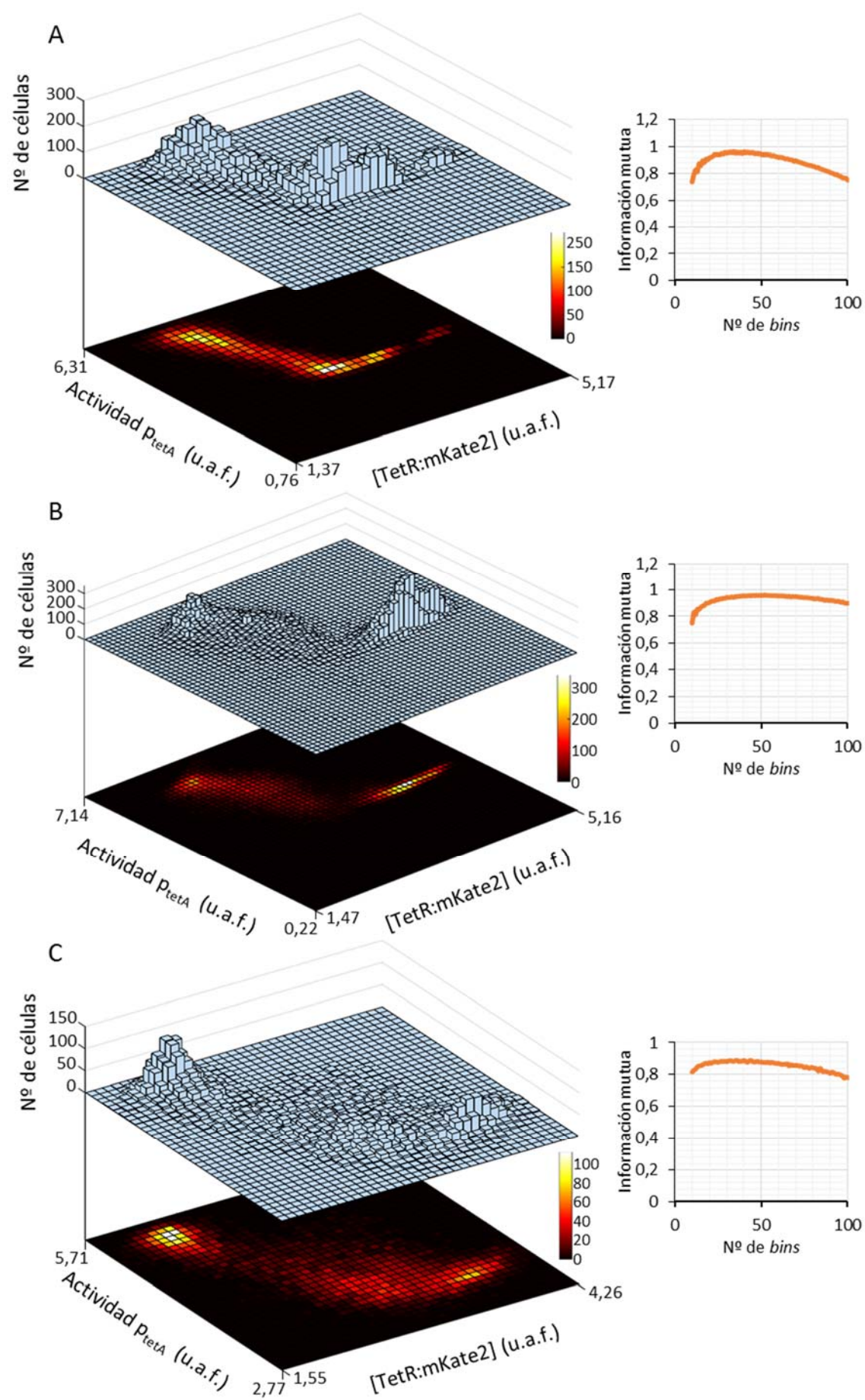
Figura 3-26. Coeficientes de variación de la actividad promotora de p_{tetA} en función de la concentración de TetR en los diferentes sistemas: (A) pRRG13+pRRG62, (B) pRRG63 y (C) RRS247. Los valores corresponden a los experimentos de las Figuras 3.16 (pRRG13+62), 3.19 (pRRG63) y 3.23 (RRS247). Cada punto corresponde a uno de los *bins* en los que se dividió la señal en el canal rojo.

3.3. Capacidad informativa del sistema regulador: cálculo de la información mutua

Como se ha observado, las diferentes células medidas presentaron valores de GFP diversos para una misma concentración de regulador, por lo tanto, la actividad transcripcional de p_{tetA} se comportó de manera estocástica. El coeficiente de variación de la actividad fue especialmente elevado en las concentraciones de TetR cercanas al valor de K de la función de transferencia. Dada esta aleatoriedad, se calculó la información mutua para medir la dependencia entre la concentración de regulador y la cantidad de GFP producida.

Como se indica en el apartado de Materiales y Métodos 2.9.3, la información mutua se calculó utilizando el método de tamizado de Grenander (Figura 3-27).

Figura 3-27. Cálculo de la información mutua mediante el método de tamizado de Grenander. Se muestra el histograma y el mapa de densidad (paneles de la izquierda) representados en la resolución óptima de *bins* por cada variable. Este número de *bins* se determinó mediante el cálculo de la información mutua máxima, que se representa en el panel de la derecha. Datos de las distribuciones de (A) pRRG13+pRRG62, (B) pRRG63 y (C) RRS247. Los mapas de densidad representan la actividad de p_{tetA} en función de la concentración de TetR:mKate2 (se muestra sólo el logaritmo en base 10 de los valores iniciales y finales de los ejes). El color del *bin* indica su densidad, creciente de negro a blanco. Los histogramas representan los mismos datos, indicando mediante la altura de la columna el número de células en cada *bin*.



Los histogramas revelaron una vez más la dispersión de la actividad promotora registrada, principalmente en el rango intermedio de concentración de regulador. Muestra de ello fue que la mayoría de las categorías representadas mostraron un número de eventos bajo o se encontraron vacías. La información mutua (I) obtenida en el caso de los tres sistemas analizados quedó cerca del bit de información. El par de vectores pRRG13 y pRRG62 obtuvo una I máxima de 0,9646 bit utilizando un tamizado de 34 *bins* para cada variable (Figura 3-27.A). pRRG63 alcanzó una I de 0,9616 bit con un tamizado de 52 *bins* (Figura 3-27.B). Por último la cepa RRS247 obtuvo una I de 0,8898 con una resolución de 36 *bins* (Figura 3-27.C). Estos valores representan el límite inferior de información transmitida, dada nuestra capacidad de detección y análisis. En los tres casos, los valores indicaron que el sistema regulador tiene una naturaleza digital (1 bit), es decir, la célula es capaz de tener activada o reprimida la producción de resistencia antibiótica en función de la señal.

Una vez calculada la información mutua para el conjunto de datos muestreados se procedió al análisis del espacio significativo. Dado que I depende de las distribuciones en las que se realiza la medida, se intentó optimizar el rango de concentración de TetR para la mejor descripción del mecanismo. Se optó por acortar la distribución por el rango del *input*, esto es, acortar por el máximo de concentración de regulador dado que era el rango en el cual la actividad no mostró cambios. De esta manera, se realizaron cálculos de I para las distribuciones decrecientes (Figura 3-28) buscando la que aportase la máxima información. El par de vectores pRRG13 y pRRG62 obtuvo una I máxima al 99,9 % de la muestra, pRRG63 la alcanzó al 99,95 % y RRS247 al acortar hasta el 98,55 %. Por lo tanto prácticamente todo el espacio muestral fue significativo. Además, observando la tendencia de las gráficas de la Figura 3-28, descendente salvo en la de RRS247, podría suponer que un incremento de muestreo en el *input* incrementaría la I conseguida en el sistema de dos vectores y de pRRG63.

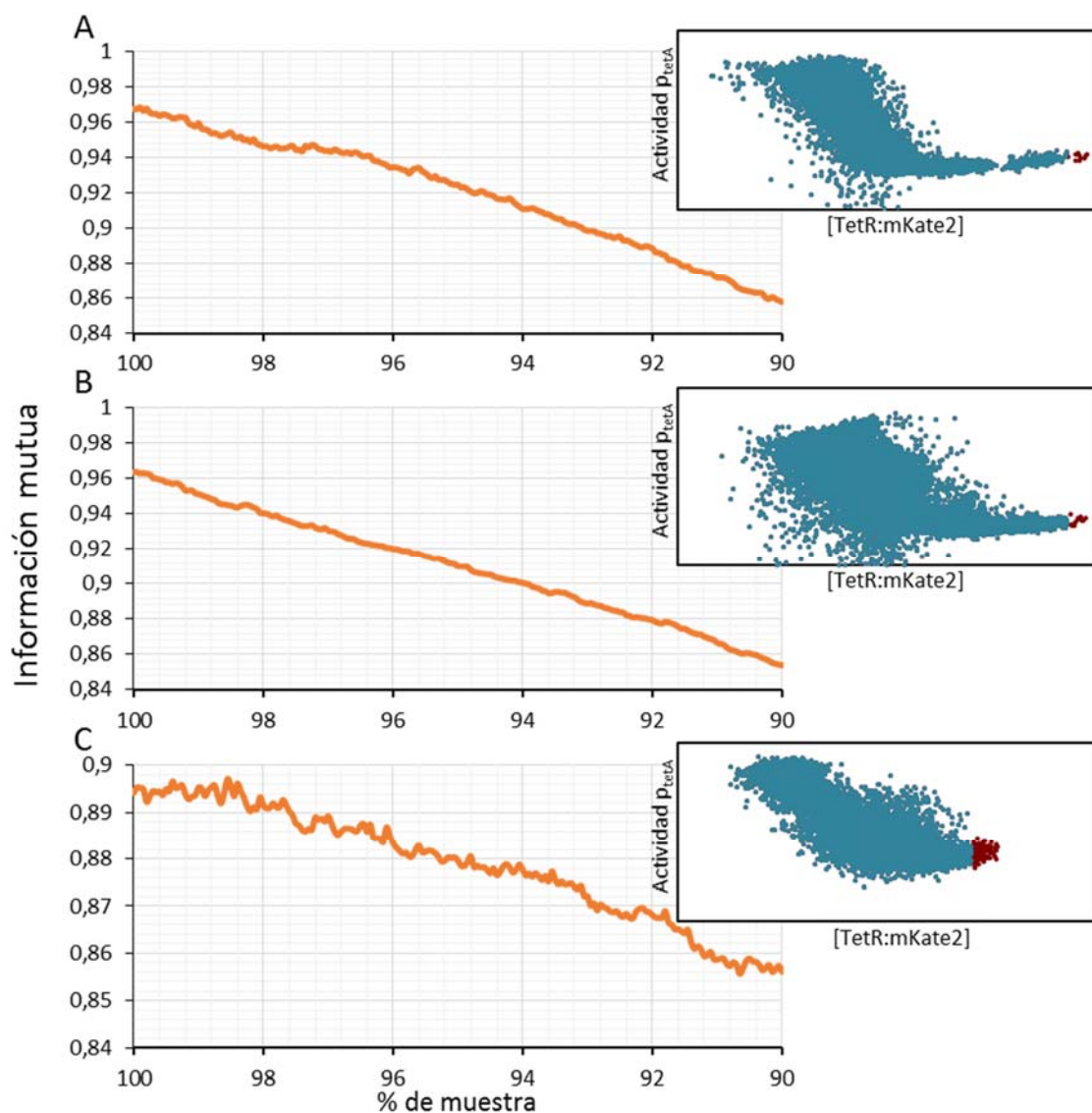


Figura 3-28. (Paneles de la izquierda) Cálculo de la información mutua para rangos decrecientes de concentración de regulador. (Paneles de la derecha) Se representa la muestra total (rojo + azul) frente a la muestra que obtuvo la información mutua máxima (azul). (A) pRRG13+pRRG62, (B) pRRG63 y (C) RRS247.

Una cuestión importante para el cálculo de la I era medir la respuesta del promotor p_{tetA} frente a todas las concentraciones fisiológicamente relevantes de regulador TetR. Para comprobar que los niveles obtenidos mediante la inducción se encontraban en el rango fisiológico, se compararon los niveles producidos por nuestras construcciones frente a los alcanzados por TetR bajo su promotor nativo. Para ello, se generó el vector pRRG74, con una región diseñada como indicadora de la actividad del operón *tet*. Es decir, en este vector, los genes *tetR:mKate2* y *gfp* quedan codificados en dirección divergente y entre ellos se localiza la secuencia intergénica *tetR-tetA* del transposón Tn10. Asimismo, se insertó esta región indicadora del vector en el cromosoma de BW27783, generando la cepa RRS252. Tanto la cepa portadora del vector como RRS252 fueron crecidas en M9 suplementado con glicerol, y al mismo tiempo se indujo el operón mediante una concentración de tetraciclina subinhibitoria del crecimiento.

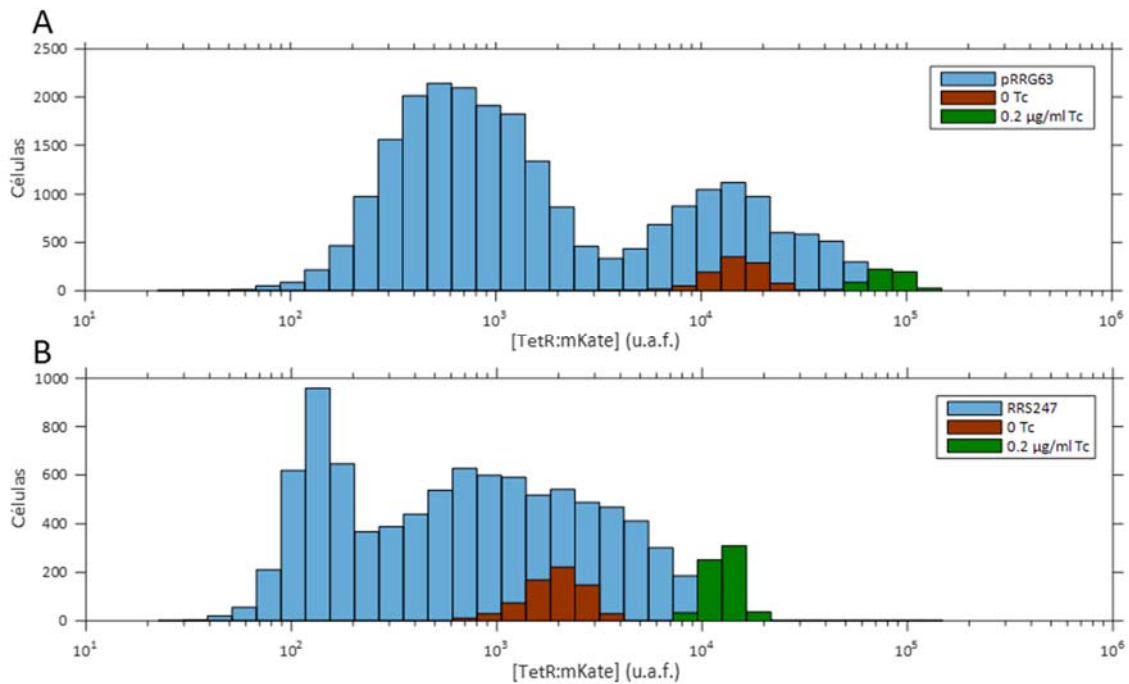


Figura 3-29. Concentración de regulador TetR producido en la configuración natural del operón *tet*. La medición se realizó como se indica en el apartado 2.8.1 de Materiales y Métodos. En verde se representan los datos de las bacterias crecidas en M9 con glicerol en presencia de 0,2µg/ml de tetraciclina y en rojo en ausencia de ella. (A) Histograma de la expresión de TetR:mKate2 en el vector pRRG74. Se representa número de células frente a concentración de TetR. Como referencia el comportamiento del vector pRRG63 figura en azul. (B) Histograma de la expresión de TetR:mKate2 en la cepa RRS252, en azul se representa el comportamiento de la cepa RRS247.

Tal y como muestra la Figura 3-29, la concentración de TetR en el estado reprimido del operón se localizó en el centro de la distribución muestreada mediante las inducciones por arabinosa, tanto para pRRG74 como para RRS252. Al inducir ambas cepas mediante 0,2 µg/ml de Tc, la expresión de TetR aumentó y alcanzó niveles marginales también representados en el muestreo realizado. No obstante, en presencia del antibiótico, TetR adquiere la conformación inactiva y no reprime el operón. Por lo tanto, estos valores enmarcan el rango fisiológico de proteína TetR activa dentro de las distribuciones de los sistemas sintéticos de medida analizados durante este trabajo, y permite presumirle una transferencia de información del mismo orden, un bit.

3.4. Análisis de los factores que modifican la información mutua en un circuito transcripcional

Una vez determinada de manera experimental la *I* entre TetR y su proteína regulada, en esta sección analizaremos de manera computacional cómo distintos factores que participan en el circuito transcripcional afectan a la capacidad informativa del canal. Para ello, utilizaremos simulaciones de Monte Carlo, que simulan la dinámica de un sistema consistente en el mRNA de la proteína reguladora, la proteína reguladora (equivalente a TetR en nuestro sistema

experimental), el mRNA de la proteína regulada y la proteína regulada (equivalente a la GFP en nuestro sistema experimental). Los métodos y algoritmos empleados para la realización de estas simulaciones se detallan en Materiales y Métodos (apartado 2.9). Hay dos factores importantes que se deben considerar a la hora de interpretar los resultados de nuestras simulaciones:

- a) Por norma general, y salvo que se especifique lo contrario, los análisis se realizaron utilizando trayectorias de 20000 iteraciones. Este número permitió una aproximación al estado estacionario, al mismo tiempo que permitió la simulación de un número suficiente de células (20000 por simulación) con un tiempo razonable de computación (entre 3 y 5 h dependiendo del sistema utilizado). Cada simulación, por tanto, contiene un total de 4×10^8 iteraciones.
- b) En nuestras simulaciones, tal y como se describe en Materiales y Métodos (apartado 2.9), las fluctuaciones en torno a la media son exclusivamente debidas a la estocasticidad en la síntesis y degradación de mRNAs y proteínas. Esto hace que el ruido observado corresponda únicamente al ruido intrínseco a la síntesis/degradación. El ruido intrínseco debido a la difusión del regulador hasta su sitio de unión, o el ruido extrínseco debido a las fluctuaciones en la maquinaria celular (RNA polimerasa, ribosomas, etc.) no aparecen reflejados en nuestros resultados.
- c) Las razones anteriormente descritas, junto al hecho de que desconocemos el número real de proteínas y mRNAs implicados (cuantificamos niveles de fluorescencia, pero no el número total de componentes), hacen que interpretemos los resultados de nuestras simulaciones de manera cualitativa. Por ejemplo, al analizar el efecto de la vida media de los componentes sobre la I , podremos establecer si la I sube o baja al incrementar o disminuir las tasas de degradación. También podremos obtener una idea aproximada de la magnitud de la dependencia, pero no podremos establecer los valores exactos para el caso de TetR/ p_{tetA} .

Pese a estas consideraciones, el análisis computacional del sistema permitió elucidar importantes principios respecto a la relación entre la función de transferencia y la capacidad informativa de las células, así como el impacto que distintos factores del proceso tienen sobre ambas.

3.4.1. La DPI en un circuito transcripcional

La primera cuestión importante a la hora de analizar nuestros resultados experimentales es determinar si un circuito transcripcional constituye un canal “clásico” de transmisión de información. Si un circuito transcripcional funcionase como un canal clásico de transmisión, la información observada entre cada uno de sus componentes para un instante t cualquiera seguiría la Desigualdad en el Procesado de la Información (DPI). La DPI indica que no es posible incrementar la cantidad de información de una señal mediante el procesamiento local, de tal manera que en una cadena de transmisión la información se pierde o se conserva en cada etapa, pero nunca incrementa. Esto significa que en nuestro sistema consistente en el mRNA del represor (x_1), la proteína represora (x_2), el mRNA del gen regulado (y_1) y la proteína correspondiente (y_2), la información entre los componentes debería seguir:

$$x_1 \rightarrow x_2 \rightarrow y_1 \rightarrow y_2$$

$$I(x_2; y_2) \leq I(x_2; y_1)$$

$$I(x_2; y_2) \leq I(y_1; y_2)$$

Ecuación 3-2. Desigualdad en el procesamiento de la información

Esto es, dado que el regulador actúa a nivel de la producción de mRNA, y este último es, a su vez, responsable de la producción de proteína regulada, la información mutua entre regulador y proteína regulada debe ser igual o menor que la I entre los componentes intermedios. Para comprobar este extremo, obtuvimos las distribuciones de los cuatro componentes para 10 niveles diferentes de inducción de TetR, simulando las trayectorias de 2000 células para cada nivel de inducción. En la Figura 3-30 se muestran las nubes de puntos correspondientes al valor en estado estacionario para las 20000 células, para trayectorias con 20000 iteraciones.

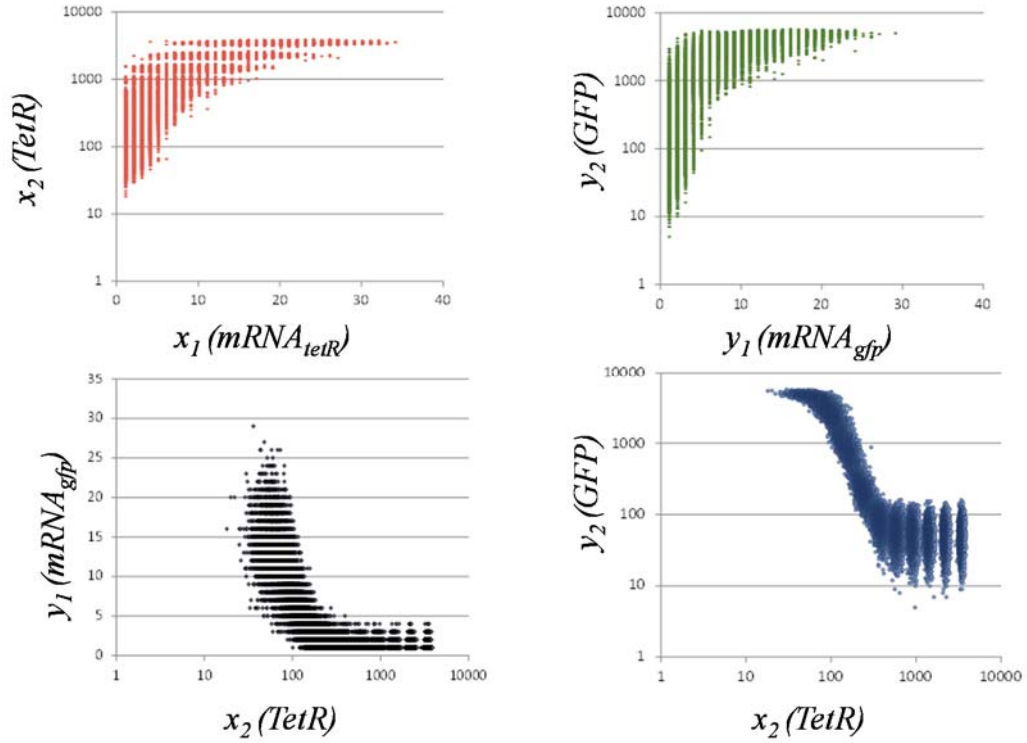


Figura 3-30. Mapa de dispersión de la simulación del circuito transcripcional. En cada uno de los paneles se representan los valores de mRNA de la proteína represora (x_1), valores de proteína represora (x_2), valores de mRNA de la proteína regulada (y_1) y valores de proteína regulada (y_2) para un total de 20000 células divididas en 10 diferentes niveles de inducción del mRNA de la proteína reguladora. Los valores representan el valor final después de calcular 20000 iteraciones para cada trayectoria.

A partir de estas distribuciones, se procedió a calcular la información mutua entre x_1 y x_2 (equivalente en nuestro diseño experimental a $I(mRNA_{tetR}; TetR)$), la información entre x_2 e y_1 $I(TetR; mRNA_{gfp})$; la información entre y_1 e y_2 $I(mRNA_{GFP}; GFP)$ y la información entre las proteínas del circuito $I(TetR; GFP)$. Para cada par de componentes se calculó I en función del número de *bins* (Figura 3-31) y se representó la distribución y las nubes de densidad de acuerdo al óptimo calculado (Figura 3-32). Como puede observarse en la Figura 3-31, para todos los *bins* analizados la cantidad de información resultó ser máxima entre TetR y GFP. Las distribuciones de la Figura 3-32 muestran que las distribuciones de mRNA y proteína se concentran en un estado densamente poblado en el que existe proteína mientras que la cantidad de mRNA es cero. Para los valores no nulos de mRNA, las distribuciones de proteína (tanto para TetR como para GFP) son altamente solapantes y, en el caso de GFP, se distribuyen casi de manera uniforme. Por el contrario, las distribuciones de TetR frente a GFP muestran un estado activado (altos niveles de GFP, bajos niveles de TetR), un estado inactivado (bajos niveles de GFP, altos niveles de TetR) y una pendiente intermedia. El valor máximo de $I(TetR; GFP)$ es de 1,6 bits, lo que correspondería a la discriminación perfecta entre 3 estados. Los mapas de densidad indican dos estados claramente diferenciables y un número

alto de estados solapantes en la pendiente intermedia entre los valores correspondientes al sistema activado e inactivado.

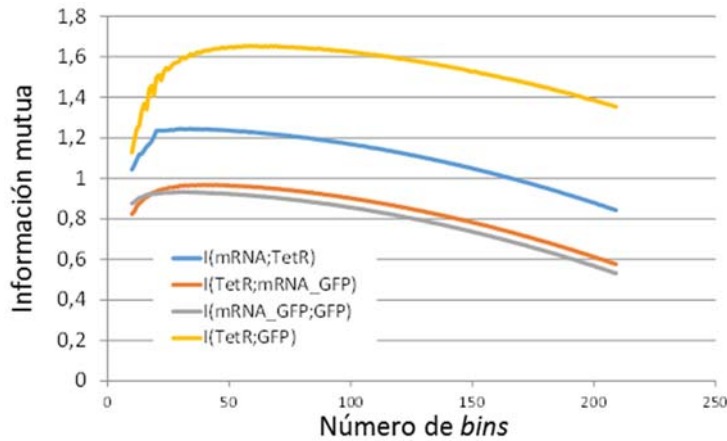


Figura 3-31. Información mutua entre cada uno de los pares de componentes del circuito transcripcional en función del número de *bins* utilizado para su cálculo. Para la distribución conjunta de cada par de componentes, se calculó la información mutua correspondiente al número de *bins* indicado en el eje de las *x*. La información mutua se computó como la suma de las entropías de las distribuciones marginales menos la entropía de la distribución conjunta $I(x,y)=H(x)+H(y)-H(x,y)$.

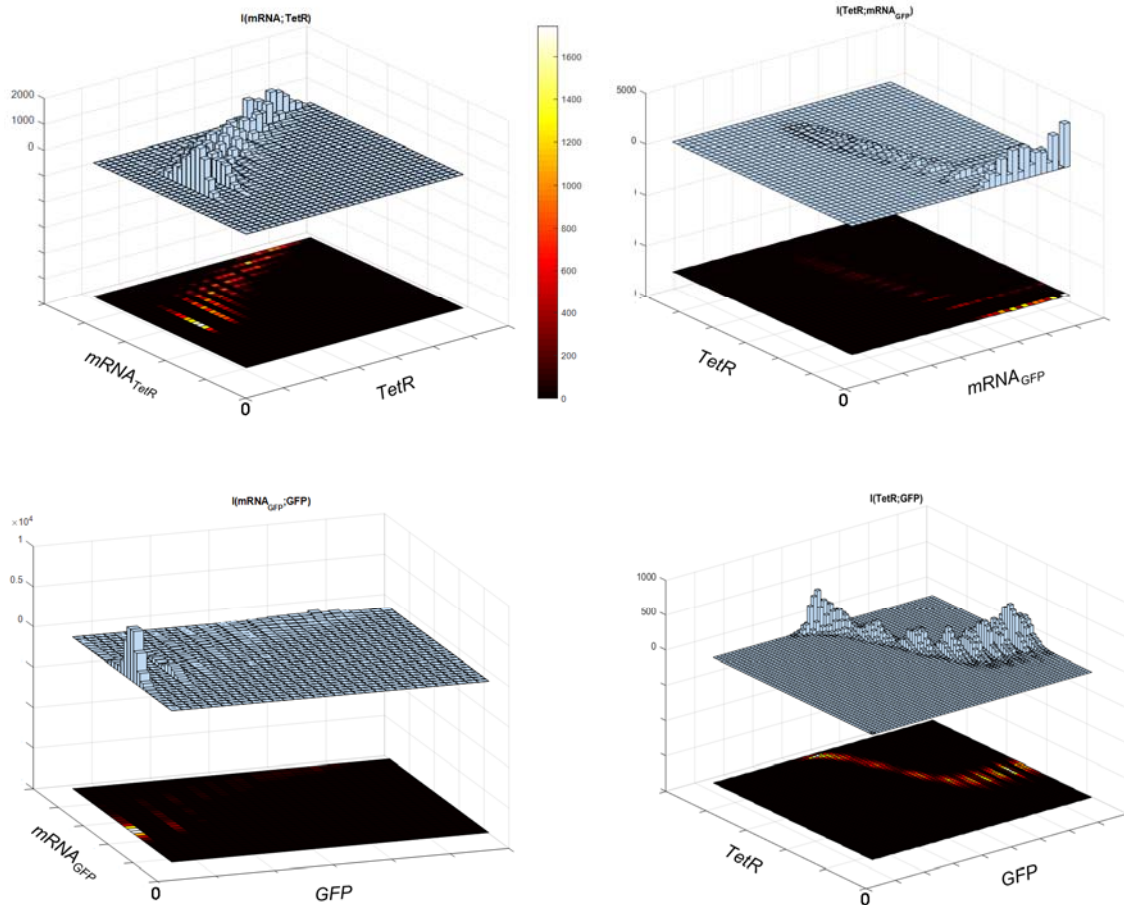


Figura 3-32. Distribuciones y mapas de densidad para los componentes del circuito. A partir de los mapas de dispersión, y basados en el número de *bins* que maximizaba la información mutua, se obtuvieron las distribuciones conjuntas para cada uno de los pares de componentes que se indican en la figura. Bajo las distribuciones se muestra el mapa de densidades, que indica el número de eventos en cada *bin*, de acuerdo a la barra de color indicada en la figura.

Los valores obtenidos para los distintos pares de componentes indicaron, inequívocamente, que $I(TetR;GFP) > I(TetR;mRNA_{gfp})$ y que $I(TetR;GFP) > I(mRNA_{gfp};GFP)$. Estos

valores, además de indicar una aparente violación de la DPI en nuestro circuito transcripcional, indicarían que la información entre las proteínas del sistema es mayor que entre estas y sus mRNAs correspondientes. Las causas y consecuencias de esta aparente violación de la DPI son analizadas en más detalle en la Discusión (apartado 4.12). Para justificar los análisis posteriores, indicaremos aquí simplemente que una diferencia fundamental entre los circuitos “clásicos” y transcripcionales radica en la vida media de los componentes que median la transmisión de la información. Mientras que en un circuito clásico la respuesta entre componentes sucede de manera casi instantánea, en un circuito transcripcional los cambios requieren una serie de reacciones químicas que influyen de manera importante en las correlaciones entre componentes. Imaginemos un interruptor conectado a una luz eléctrica. En dicho circuito, la observación de uno de los dos componentes en un punto dado informa, inequívocamente, del estado del otro componente. Esto es así porque la interrupción de la corriente apaga la luz de manera instantánea. Por el contrario, imaginemos un circuito transcripcional similar, en el que la producción de una proteína requiere de la presencia de un activador transcripcional. Incluso aunque el activador funcionase de manera perfectamente digital, y su desaparición fuese instantánea, la ausencia de dicho activador no se traduce de manera inmediata en la ausencia de proteína diana, puesto que la degradación de la proteína diana requiere cierto tiempo (Figura 3-33).

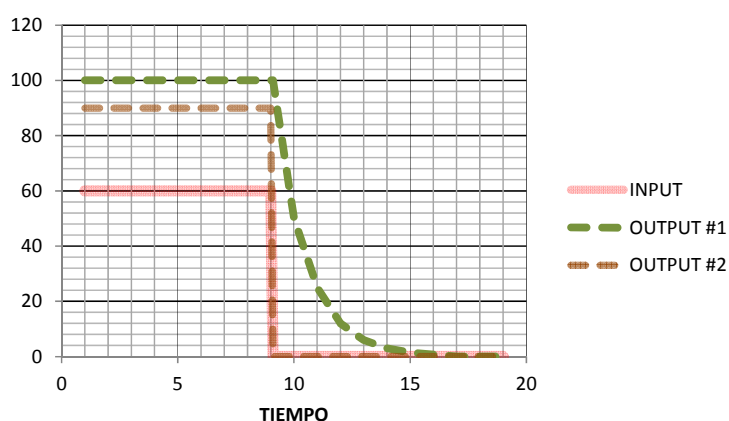


Figura 3-33. Evolución en el tiempo de un *output* de respuesta instantánea (*OUTPUT #2*) y un *output* de respuesta temporal (*OUTPUT #1*) frente a un regulador digital de respuesta instantánea (*INPUT*).

Como consecuencia, es posible que las vidas medias de los componentes tengan un impacto importante en la información mutua entre dos componentes de un circuito transcripcional.

3.4.2. Influencia de la vida media de los componentes en la información mutua.

Con el objeto de estudiar el efecto de la vida media en la capacidad informativa de un canal transcripcional, realizamos simulaciones en las que variamos sistemáticamente la tasa de degradación de las proteínas del sistema. En bacterias, mRNA y proteínas tienen vidas medias

características radicalmente distintas. Mientras que los mRNAs bacterianos tienen vidas medias cortas de alrededor de 1-10 min, la vida media de las proteínas es variable, pero se encuentra típicamente varios órdenes de magnitud por encima (Larrabee *et al.*, 1980). En nuestro circuito transcripcional experimental, las proteínas fluorescentes presentes tanto en el regulador transcripcional como en la proteína regulada pueden asumirse como estables, con una vida media superior a 1 día para GFP (Andersen *et al.*, 1998). Dado que no conocemos las vidas medias de TetR nativo ni de su proteína regulada (TetA), decidimos estudiar si un cambio en la vida media de estas proteínas afectaría a la I mostrada por el canal. Para ello, definimos E como el cociente entre la tasa de degradación de las proteínas del sistema, dividida por la tasa de degradación de los mRNAs (que fijamos en 3 minutos). Realizamos simulaciones para $E=100$ (proteínas altamente estables), $E=10$ (proteínas relativamente estables) y $E=1$ (proteínas inestables). Para cada condición se simuló la trayectoria de un total de 20000 células, divididas en 10 niveles diferentes de inducción de regulador, y se obtuvieron los valores correspondientes para cada especie tras 20000 iteraciones. En la Figura 3-34 se muestran los resultados para el regulador (X , TetR) y proteína regulada (Y , GFP). En la Figura 3-35 se muestran las distribuciones y la I entre el mRNA del regulador (x , mRNA_{TetR}) y el regulador (X , TetR); mientras que en la Figura 3-36 se muestra las distribuciones y la I entre el regulador (X , TetR) y el mRNA de la proteína regulada (y , mRNA_{GFP}).

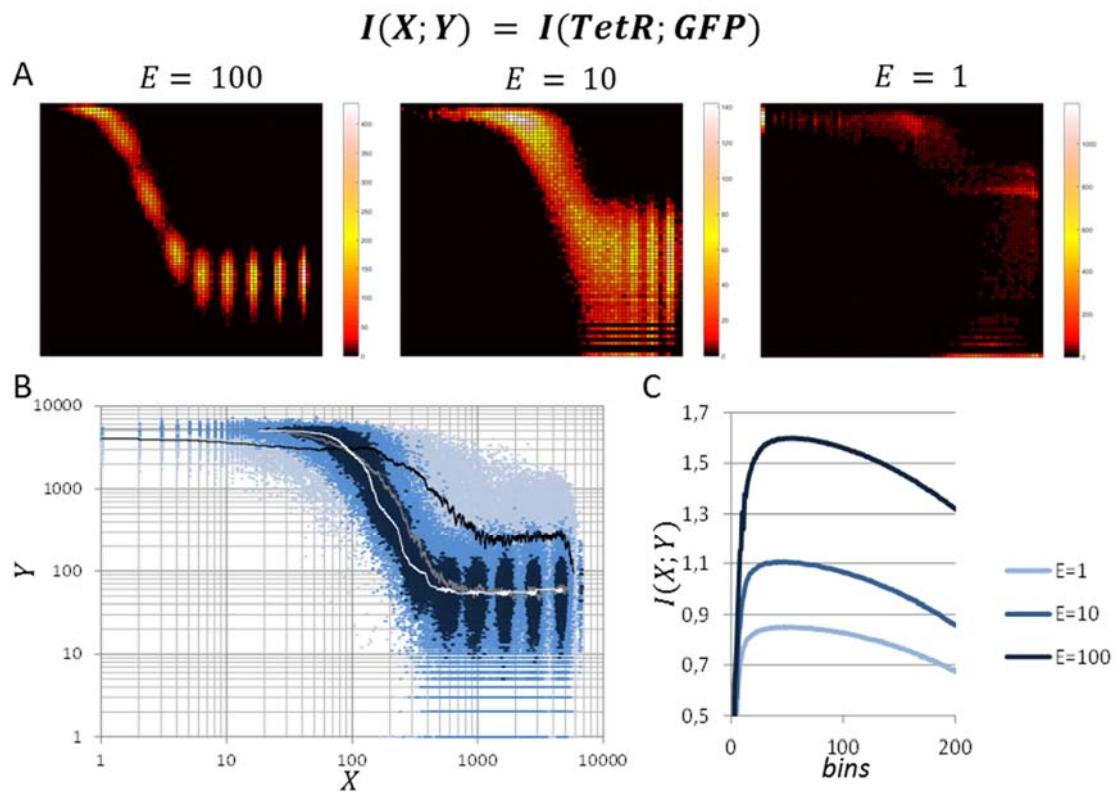


Figura 3-34. Efecto del cambio en la vida media de las proteínas en la I entre TetR y GFP. Se realizaron simulaciones modificando E (tasa de degradación de las proteínas dividida entre la tasa de degradación del RNA). (A) Mapas de densidad dependiendo de E . En el eje vertical se representa TetR (X) frente a GFP (Y) en el eje horizontal. Los colores indican mayor densidad, de negro a blanco. (B) Nubes de puntos de los datos simulados. Se representa el valor de Y para cada célula en función de X . (C) Cambio del valor de la información mutua en función del número de *bins* utilizados para su cálculo. Los colores en los paneles B y C indican el valor de E en cada simulación según el código en C.

Como puede observarse en la Figura 3-34, al disminuir la vida media de las proteínas ($E \rightarrow 1$), la I entre regulador y proteína regulada descendió significativamente, desde valores de $I(X; Y) = 1,6$ para $E = 100$, hasta $I(X; Y) = 0,8$ para $E = 1$. Estos resultados indican que un descenso en la estabilidad de las proteínas conllevaría una disminución en la I observable entre las distribuciones de ambas. Como apuntamos antes, las proteínas fluorescentes son esencialmente estables (Andersen *et al.*, 1998), por lo que, aunque desconocidas, las vidas medias de TetR nativo y TetA pueden asumirse como similares o en todo caso inferiores a las de las variantes fluorescentes. Esto indica que la I calculada experimentalmente debe considerarse un límite superior a la I del sistema no marcado.

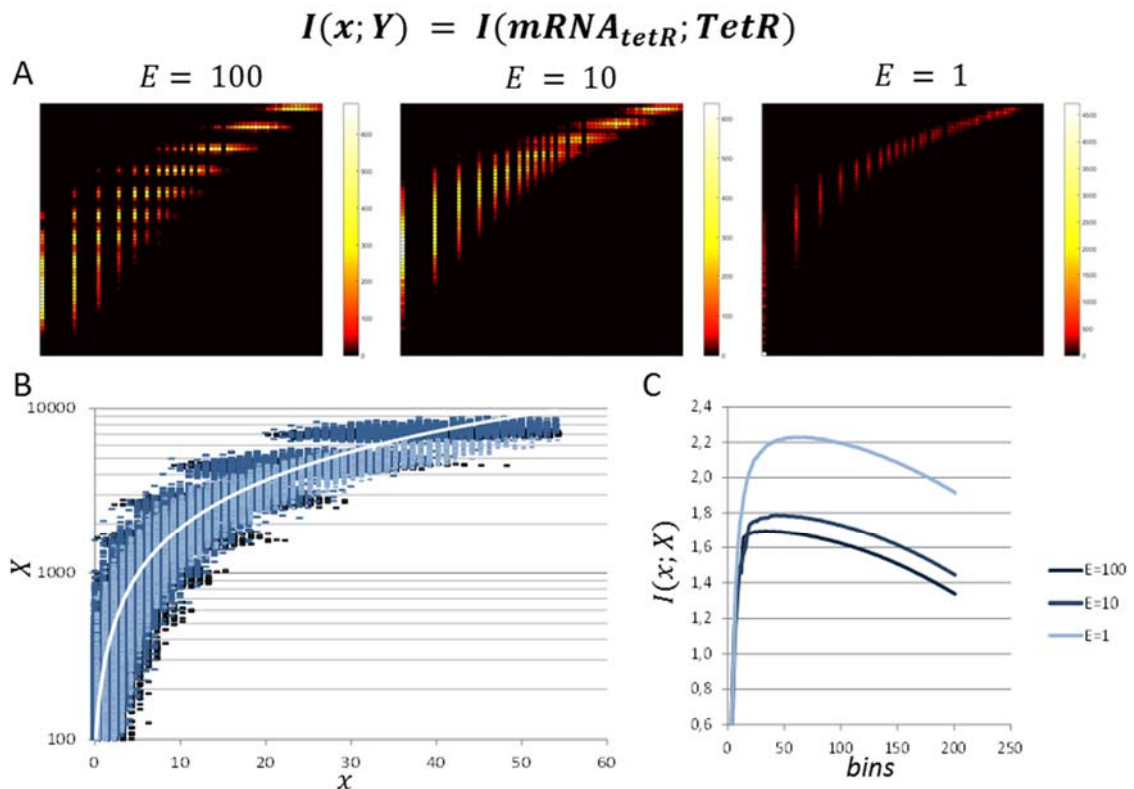


Figura 3-35. Efecto del cambio en la vida media de las proteínas en la I entre $mRNA_{tetR}$ y TetR. Se realizaron simulaciones modificando E (tasa de degradación de las proteínas dividida entre la tasa de degradación del RNA). (A) Mapas de densidad. En el eje vertical se representa TetR (X) frente a $mRNA_{tetR}$ (x) en el eje horizontal. Los colores indican mayor densidad, de negro a blanco. (B) Nubes de puntos de los datos simulados. Se representa el valor de X para cada célula en función de x . (C) Cambio de la información mutua en función del número de *bins* utilizados para su cálculo. Los colores en los paneles B y C indican el valor de E según el código en C.

La Figura 3-35 indicó, no obstante, que a medida que la estabilidad de la proteína se acerca a la de su mRNA, la I entre ellos aumenta, alcanzándose un máximo de $I(x; X)=2,2$ cuando $E=1$. Este resultado nos muestra que a) un mRNA y su proteína correspondiente muestran mayores grados de correlación cuanto más similares son sus tasas de degradación y b) esto no indica, sin embargo, que una mejor correlación entre mRNA y proteína resulte en una mayor I entre proteína reguladora y proteína regulada.

Por último, en la Figura 3-36 se muestra que la I entre el regulador y el mRNA regulado no cambia significativamente ante variaciones en E . Esto es debido a que, aunque la información es mutua, la dependencia de la I respecto a ambas variables no es equivalente. En el caso de la relación entre el regulador y el mRNA regulado, este último se asume como inestable, por lo que su capacidad de “memorizar” las concentraciones pasadas de regulador es mínima. Es por esto que la I se hace relativamente independiente de la vida media del regulador: sea cual sea su concentración y su propia “memoria”, el mRNA regulado solo es capaz de contener la I que su vida media permite.

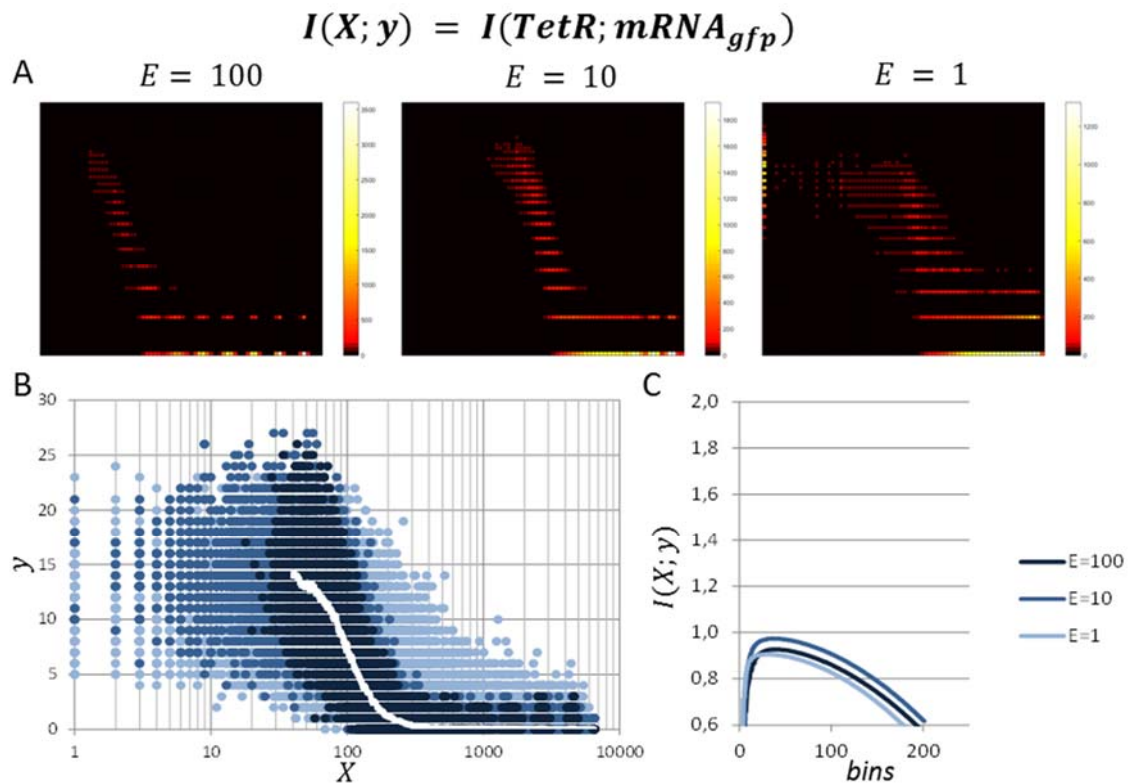


Figura 3-36. Efecto del cambio en la vida media de las proteínas en la I entre TetR y mRNA_{gfp}. Para producir los distintos comportamientos se realizaron simulaciones modificando E (tasa de degradación de las proteínas dividida entre la tasa de degradación del RNA). (A) Mapas de densidad, en las abscisas se representa TetR (X) y en el eje de ordenadas mRNA_{gfp} (y). Los colores indican mayor densidad, de negro a blanco. (B) Nubes de puntos de los datos simulados. Se representa el valor de y para cada célula en función de X . (C) Cambio de la información mutua en función del número de bins utilizados para su cálculo. Los colores en los paneles B y C indican el valor de E según el código en C.

3.4.3. Efecto de las fluctuaciones sobre la información mutua.

Además de la vida media de los componentes, otro factor importante que puede afectar a la capacidad de un circuito transcripcional para transmitir información es el nivel de fluctuaciones experimentadas por sus componentes. Dado que las reacciones químicas de síntesis y degradación de RNA y proteínas son intrínsecamente estocásticas, la célula mantiene un control imperfecto sobre las concentraciones de los componentes del circuito (Lestas *et al.*, 2010). Esta aleatoriedad en las reacciones de síntesis y degradación se traduce en la existencia de un cierto "ruido" genético, tanto en la señal (el TF), como en el *output* (la proteína regulada). En general se acepta que, como en cualquier otro canal de información, el ruido disminuye la capacidad informativa. Sin embargo, cuantificar el impacto del ruido sobre la información mutua es complicado. No disponemos de métodos experimentales adecuados que nos permitan alterar específicamente las fluctuaciones de los componentes, lo que dificulta un abordaje experimental directo. Por otra parte, aunque existen soluciones analíticas que correlacionan el nivel de ruido con la capacidad informativa de un factor de transcripción, dichas soluciones se basan en aproximaciones gaussianas (Tkacik *et al.*, 2008). Desgraciadamente, este tipo de aproximaciones (conocidas como *small noise approximations*) resulta poco adecuado para el problema que nos ocupa. En bacterias, los reguladores transcripcionales se encuentran en bajo número de copias (Hammar *et al.*, 2012; Taniguchi *et al.*, 2010), dando lugar a distribuciones poissonianas (Paulsson, 2004) o binomiales negativas (So *et al.*, 2011). Otras soluciones, basadas en relaciones de fluctuación-disipación, si bien no contienen asunciones previas sobre la distribución de componentes, requieren aproximaciones lineales al ruido (Paulsson, 2004), poco adecuadas para sistemas como TetR, que mostró índices de cooperatividad cercanos a 4. Por estas razones, decidimos abordar el estudio del impacto de las fluctuaciones moleculares sobre la capacidad informativa del canal de manera computacional.

Para ello, realizamos simulaciones en las que mantuvimos constantes las tasas de síntesis y degradación de los componentes, así como los parámetros "mecánicos" del sistema (la constante K y el grado de cooperatividad n). Sin embargo, modificamos sistemáticamente el tamaño de la salva S (*burst size*) en las reacciones de síntesis. Este es el principal factor generador de ruido intrínseco (So *et al.*, 2011), y se corresponde al número promedio de moléculas producidas por reacción de transcripción/traducción. Cuando $S=1$ el sistema se encuentra en el límite poissoniano de fluctuaciones, mientras que para $S>1$ las distribuciones toman forma de binomiales negativas con mayores niveles de ruido (So *et al.*, 2011). Para estudiar el impacto del ruido intrínseco sobre la capacidad informativa del canal, analizamos

de manera independiente el efecto de distintos niveles de S en la traducción de TetR (*input*) y de GFP (*output*).

Efecto de las fluctuaciones intrínsecas sobre el input

Para analizar el efecto de las fluctuaciones intrínsecas en TetR, computamos las trayectorias de 20000 células, sometidas a 10 condiciones de inducción diferentes, utilizando el código para poblaciones indicado en el apartado 2.9.2 de Materiales y Métodos. De manera análoga a lo indicado anteriormente, para cada célula se recuperaron las concentraciones en estado estacionario de mRNAs y proteínas después de un total de 20000 iteraciones. Repetimos este análisis para valores de S entre 1 y 5, y de las distribuciones obtenidas se computó el nivel de ruido intrínseco en la señal (TetR), el mapa de respuesta TetR/GFP y la información mutua observable. Los resultados del análisis se muestran en la Figura 3-37. Como puede observarse en el panel C de la figura, a medida que aumentamos el tamaño de S , los niveles de ruido en TetR aumentan. Al computar las distribuciones y la información mutua observable, comprobamos que este aumento en el nivel de ruido intrínseco en el *input* se traduce en una reducción en la capacidad informativa del canal, que pasa de 1,8 bits en el límite poissoniano a 1,3 bits cuando $S=5$ (Figura 3-37.D). Este cambio en I se produce sin que el valor promedio de la función de transferencia sufra alteraciones importantes (Figura 3-37.C). El único cambio aparente es un incremento en el rango de X , debido a que un aumento en S_X produce un mayor número promedio de proteínas por mRNA, por lo que $S=5$ alcanza concentraciones de X 5 veces mayores que $S=1$. Dado que el rango de X es un factor crítico a la hora de determinar la I , nos preguntamos si las $I(X;Y)$ observadas entre los distintos valores de S se debían exclusivamente al cambio en la distribución marginal de X . Para comprobar este extremo, fijamos un máximo de X , X_{max} y nos quedamos con los pares XY para todos los valores de S en los que $X < X_{max}$. Para cada valor de S se eligieron aleatoriamente 15000 pares XY y se computó $I(X;Y)$. Este tratamiento normaliza tanto el rango como el número de observaciones para todos los valores de S , lo que permite determinar si el incremento del ruido intrínseco en la señal produce algún efecto sobre $I(X;Y)$ independiente de la amplitud de la distribución marginal de X . Como puede verse en el panel E de la Figura 3-37, un aumento de S produce un descenso en $I(X;Y)$ incluso cuando las distribuciones marginales son iguales. De estos resultados concluimos que, para una función de transferencia dada, un aumento en el ruido intrínseco produce una disminución en la capacidad informativa del canal.

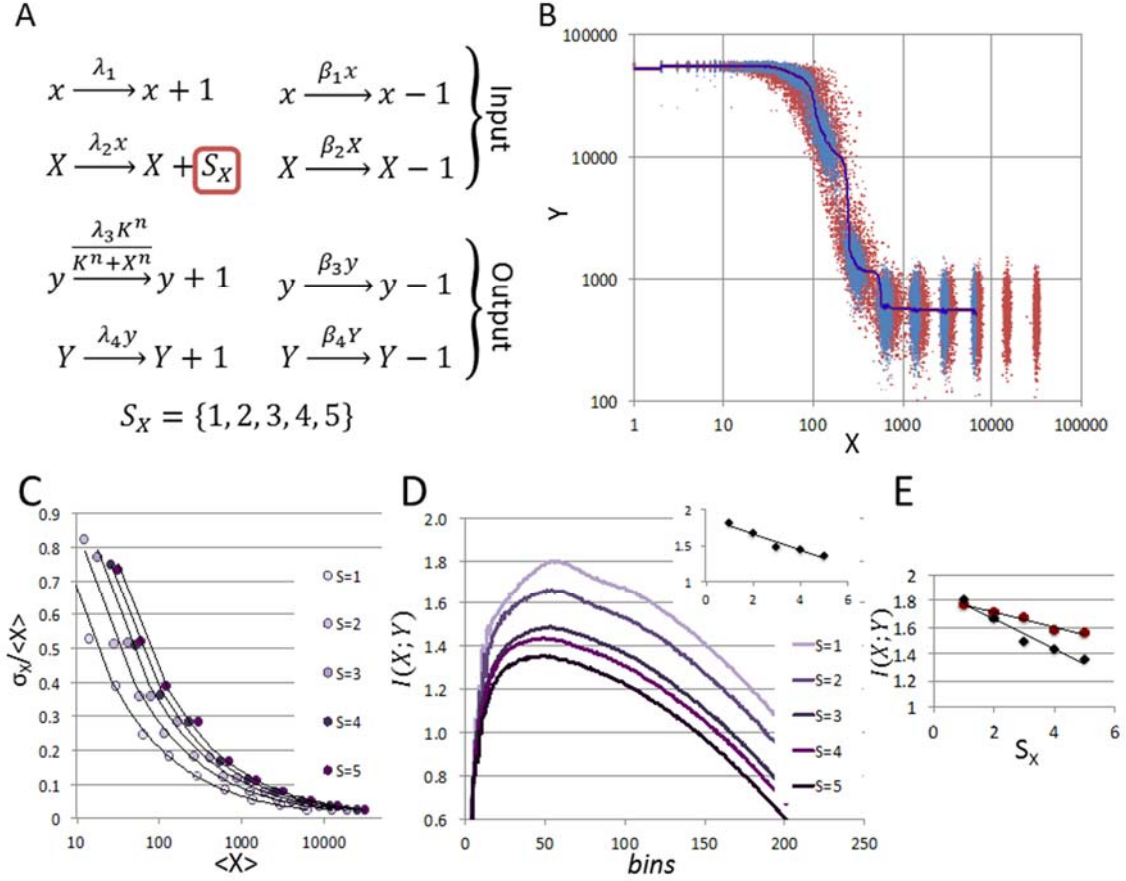


Figura 3-37. Efecto de las fluctuaciones en el *input* sobre la capacidad informativa del canal. (A) CMEs del sistema. Para producir un incremento en las fluctuaciones del *input* modificamos el tamaño de la salva, S_X , en la producción de X . Los resultados de la figura indican valores desde $S=1$ proteína por evento de traducción, hasta $S=5$ proteínas por evento de traducción. (B) Mapa de *input* (X)/*output* (Y) para $S=1$ (puntos azules) y $S=5$ (puntos rojos). Las líneas continuas indican la media móvil para ambas poblaciones (roja para $S=5$ u azul para $S=1$). Ambas líneas aparecen solapantes. (C) Nivel de fluctuaciones para distintas concentraciones promedio de X . Se muestra en el panel los valores de fluctuación expresados como coeficiente de variación frente al valor promedio de x para distintos valores de S . (D) Información mutua $I(X;Y)$ calculada con distinto número de bins (eje x) para los distintos valores de S . En el recuadro interno se muestra el valor máximo de $I(X;Y)$ obtenido en función de S . (E) Valor máximo de I para distintos valores de S cuando se permiten distintas distribuciones marginales de X (puntos negros) y cuando I es calculada sobre una misma distribución marginal de X (puntos rojos).

Efecto de las fluctuaciones intrínsecas sobre el output

De igual forma, analizamos el impacto sobre la capacidad informativa del canal de cambios en el tamaño de la salva para la producción del *output*, GFP. Las simulaciones se realizaron de manera idéntica a lo descrito anteriormente, incrementando el valor de S para la traducción de la GFP de 1 molécula a 5. A partir de las trayectorias se obtuvo la distribución en estado estacionario y de estas distribuciones se calculó el valor máximo de $I(TetR; GFP)$. Como puede verse en la Figura 3-38 existen diferencias significativas entre estos resultados y los obtenidos para el *input* (Figura 3-37). En primer lugar al modificar S_Y se modifica el rango de Y de manera acorde (Figura 3-38.B). Esto hace que la función de transferencia promedio sea diferente (para un mismo valor de X se obtienen distintos valores de Y). Sin embargo, al

calcular $I(X, Y)$ observamos que un aumento en S_Y no produjo diferencias significativas en la cantidad de información transmitida por el canal (Figura 3-38.D). De hecho, al representar las fluctuaciones en el *output* frente a su valor medio (Figura 3-38.C) observamos que para un amplio rango de valores de Y , el aumento en S_Y no produce variaciones apreciables en el ruido. El resultado es que, al aumentar el valor de S_Y se produce un traslado de la función de transferencia hacia valores más altos de Y . Sin embargo, debido a que las fluctuaciones en Y provienen fundamentalmente de los niveles de *input*, este traslado no varía significativamente las fluctuaciones en el *output*. Como consecuencia, $I(X; Y)$ para todos los valores de S_Y fueron similares. De estos resultados podemos concluir que: a) una modificación en el rango de Y puede alterar la función de transferencia sin alterar su capacidad de transmisión de información; y b) variaciones en el ruido intrínseco del *output* tuvieron un efecto mínimo sobre la capacidad informativa del canal.

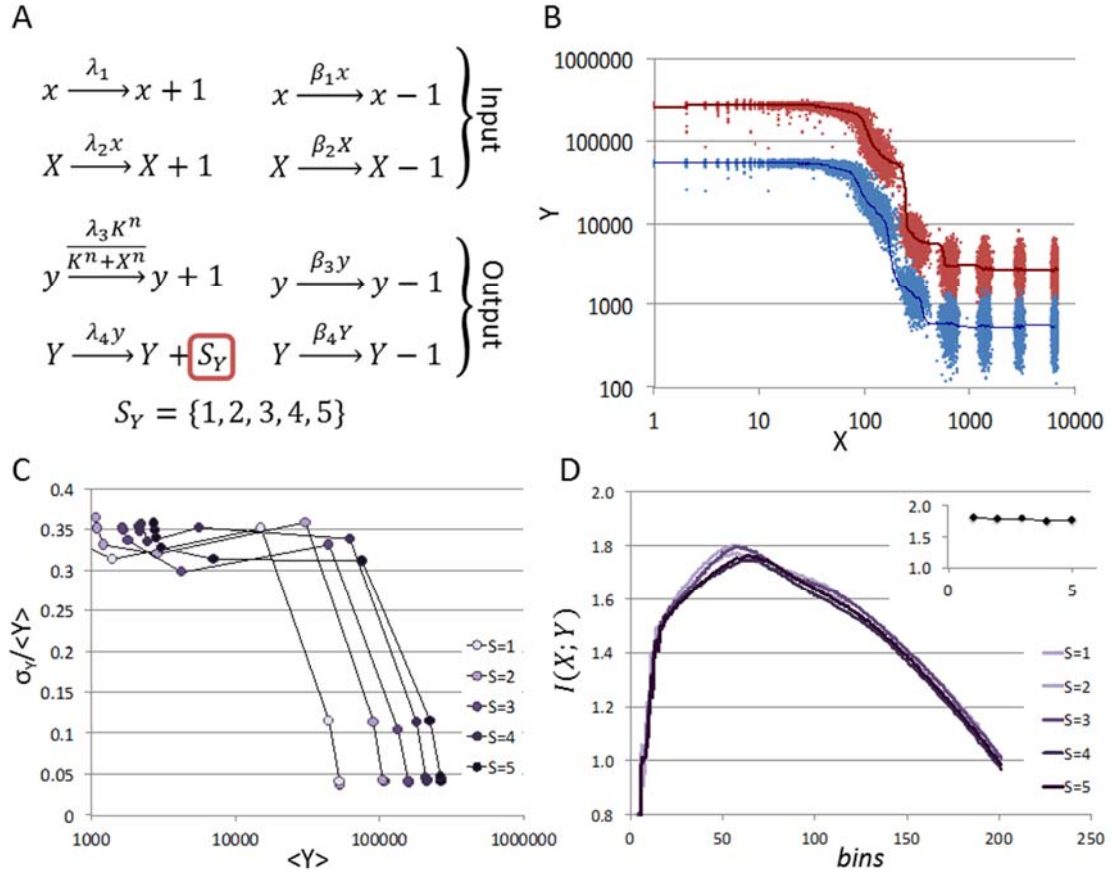


Figura 3-38. Efecto de las fluctuaciones en el *output* sobre la capacidad informativa del canal. (A) CMEs del sistema. Para producir un incremento en las fluctuaciones del *output* modificamos el tamaño de la salva, S_Y , en la producción de Y . Los resultados de la figura indican valores desde $S=1$ proteína por evento de traducción, hasta $S=5$ proteínas por evento de traducción. (B) Mapa de *input* (X)/*output* (Y) para $S=1$ (puntos azules) y $S=5$ (puntos rojos). Las líneas continuas indican la media móvil para ambas poblaciones (roja para $S=5$ y azul para $S=1$). (C) Nivel de fluctuaciones para distintas concentraciones promedio de Y . Se muestra en el panel los valores de fluctuación expresados como coeficiente de variación frente al valor promedio de Y para distintos valores de S . (D) Información mutua $I(X; Y)$ calculada con distinto número de *bins* (eje x) para los distintos valores de S . En el recuadro interno se muestra el valor máximo de $I(X, Y)$ obtenido en función de S .

3.4.4. Efecto de K y n sobre la información mutua

La razón fundamental de que un TF/gen regulado mantenga un cierto nivel de I es que existe una función de transferencia que correlaciona la producción del gen regulado con los niveles de TF. Es por esto que, además de factores propios de los componentes como la vida media y los niveles de fluctuación, los parámetros que modifican la forma de la función de transferencia también tienen un impacto directo sobre la capacidad informativa del canal. Tkačik, Walczak y Bialek (Tkacik *et al.*, 2009) estudiaron el problema de, dado un rango de concentraciones de *input* y *output*, y unos niveles de ruido asociados a ambas variables, qué función de transferencia optimiza la I . Para encontrar una constante K (constante de disociación de la unión del regulador a su operador) y un grado de cooperatividad n óptimos, Tkačik *et al.* recurrieron a una aproximación para el ruido de entrada basada en que este era de pequeña magnitud y normalmente distribuido. Asumiendo distribuciones gaussianas en el *input*, pudieron obtener expresiones analíticas para la I , que posteriormente optimizaron en función de K y n . Nuestras simulaciones con CMEs no asumen a priori que la distribución de componentes deba seguir una forma concreta. No obstante, dado que asumimos tiempos de espera exponenciales, los componentes tienden a distribuirse en forma binomial o poissoniana. Para comprobar si estos cambios en la distribución de fluctuaciones hacían cambiar los óptimos obtenidos por Tkačik *et al.* para distribuciones normales, realizamos una serie de simulaciones en las que modificamos los valores de K y de n .

Las simulaciones se realizaron de manera idéntica a lo descrito anteriormente (apartado 2.9.2 de Materiales y Métodos), haciendo variar el valor de cooperatividad aparente $n=\{1,2,4,6,8\}$. Los resultados se muestran en la Figura 3-39. Los cambios en n produjeron, no solo una modificación en la forma de la función de transferencia, sino una redistribución de las masas de probabilidad. Así para $n=1$ y $n=2$ las células pueblan de manera más o menos uniforme toda la trayectoria de la función. Sin embargo, al aumentar n , hay una concentración de las masas de probabilidad sobre los extremos de la función. Un aumento de la n , por tanto, no solo produce una pendiente más pronunciada alrededor de K , sino que disminuye la proporción de células que se encuentran en estados intermedios. En el límite $n \rightarrow \infty$ esperaríamos, por tanto, un interruptor digital perfecto en el que todas las células estuviesen en uno de los dos estados posibles. Como era de esperar, esto se correlacionó con una tendencia a que valores altos de n disminuyesen la I . No obstante, esta relación no es monótona, dado que el máximo de I se observó para $n=2$. Este resultado es similar al resultado analítico obtenido por Tkačik para fluctuaciones pequeñas y normales.

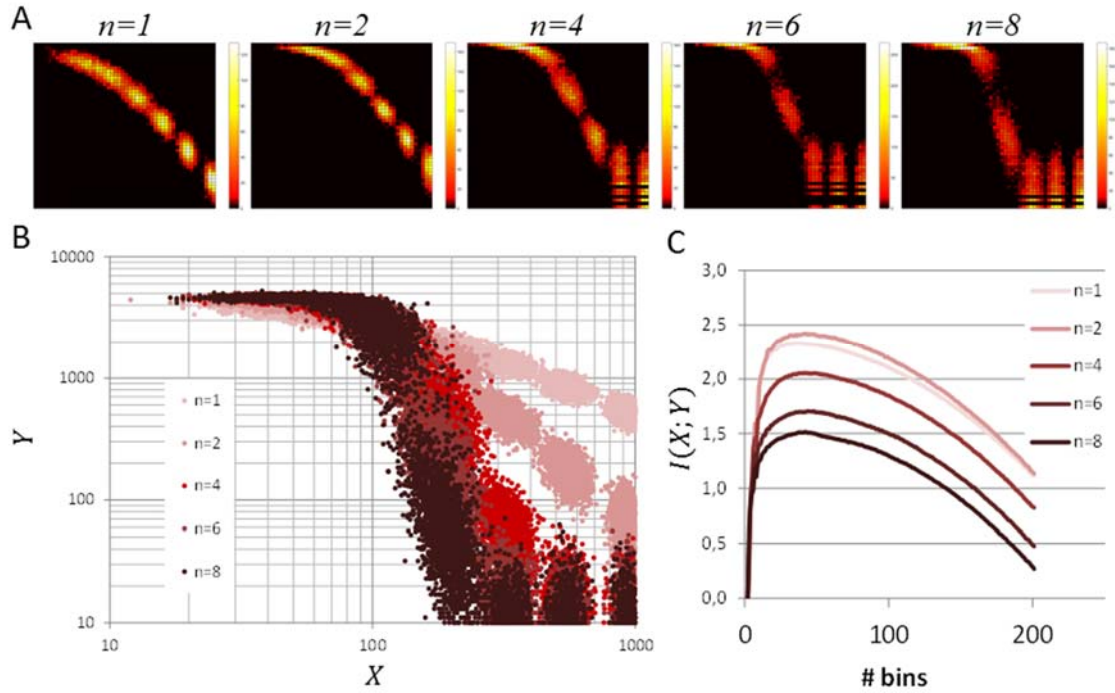


Figura 3-39. Efecto de n sobre la información mutua. Para producir los distintos comportamientos se realizaron simulaciones modificando el valor de n . (A) Mapas de densidad para las distintas n (cada punto representa una célula). Se representa actividad promotora (Y) frente a regulador (X). Los colores indican incremento en la densidad, de negro a blanco. (B) Nubes de puntos de los datos simulados. Se representa el valor de Y para cada célula en función de X . Los colores indican el valor de n según la leyenda del gráfico. (C) Cambio de la información mutua en función del número de *bins* para las diferentes n .

De forma análoga, fijamos $n=4$ (correspondiente a los valores aproximados obtenidos experimentalmente) y variamos sistemáticamente K . Los resultados, mostrados en la Figura 3-40, indican la relación entre el valor de K/c , donde c indica la concentración máxima de regulador, y el valor de la I entre el regulador X y su proteína regulada Y . Como puede observarse en el panel C de la Figura 3-40, la relación entre K/c y la I no es monótona, alcanzándose un máximo en $K=0,1c$. Valores de K cercanos a 0 sufren del problema del enorme efecto de las fluctuaciones intrínsecas para valores bajos de X . Por el contrario, altos valores de K hacen que la mayor parte del rango de X no produzca cambios aparentes en Y , por lo que la información mutua disminuye.

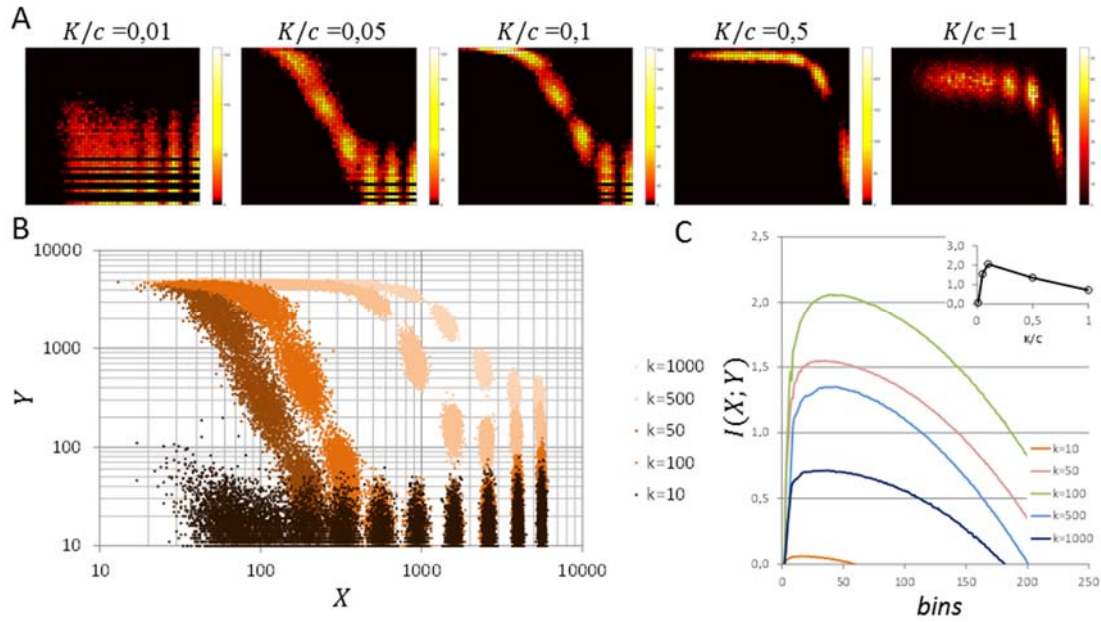


Figura 3-40. Efecto de K sobre la información mutua. Para producir los distintos comportamientos se realizaron simulaciones modificando el valor de K . (A) Mapas de densidad de las diferentes distribuciones (cada punto representa una célula) dependiendo del ratio K/c (constante de disociación entre concentración máxima de regulador). Se representa actividad promotora (Y) frente a regulador (X). Los colores indican incremento en la densidad, de negro a blanco. (B) Nubes de puntos de los datos simulados. Se representa el valor de Y para cada célula en función de X . Los colores indican el valor de K . (C) Cambio de la información mutua en función del número de *bins* para las diferentes K . En el recuadro se representa la I_{max} obtenida respecto a K/c .

Discusión

4. Discusión

4.1. El problema de la parametrización de las redes de regulación transcripcionales

La transcripción de los genes, como ya hemos visto, está regulada por proteínas llamadas factores de transcripción. Estos factores a su vez están codificados en otros genes cuya transcripción está controlada por otros TF (o por ellos mismos), y así sucesivamente, generando una red de vínculos de regulación entre ellas. Además, los factores de transcripción pueden ver modificada su actividad por acción de otras moléculas. Al conjunto de estas interacciones y las partes participantes las llamamos redes de regulación. El estudio clásico de estas redes se ha basado en el conocimiento de su topología, es decir, qué componentes conforman la red y qué interacciones hay entre ellos. La topología es un factor crítico en el comportamiento de las redes transcripcionales, pero igualmente importante es cuantificar las interacciones entre los nodos, lo que denominamos problema paramétrico (Fernandez-Lopez *et al.*, 2010; Rosenfeld *et al.*, 2005). Para resolver este problema es necesario cuantificar la relación entre los nodos, lo que en redes transcripcionales implica la cuantificación de la relación entre un TF y su promotor diana. Esta función se denomina función de transferencia. Obtener esta función plantea una serie de problemas metodológicos y teóricos:

- i) ¿Qué métodos experimentales son los adecuados para su obtención? ¿Pueden sistematizarse?
- ii) ¿Qué nivel de descripción es necesario para poder obtener funciones que expliquen el comportamiento del sistema y que permitan predecir su comportamiento? ¿Es posible obtenerla mediante medidas poblacionales o, por el contrario, requiere medidas en células individuales?
- iii) ¿Qué unidades podemos utilizar para describirla? ¿Qué modelos matemáticos permiten una mejor aproximación al problema de la transducción de señales en las redes transcripcionales?

Partiendo de diferentes aproximaciones a este problema (Fernandez-Lopez *et al.*, 2010; Rosenfeld *et al.*, 2005; Tkacik *et al.*, 2008) en esta tesis, hemos explorado distintas metodologías para dar respuesta a estas preguntas.

4.2. La fluorimetría de cultivos presenta problemas de sensibilidad a la hora de determinar la función de transferencia.

Desde su introducción, la fluorimetría de cultivos es una técnica relativamente sencilla, ampliamente utilizada para caracterizar la respuesta dinámica de un promotor (Kalir *et al.*, 2001; Ronen *et al.*, 2002). Esta técnica permite extraer valores cuantitativos sobre la expresión desde un promotor determinado, utilizando proteínas fluorescentes como genes reporteros. Así, la medida de los valores de fluorescencia y de densidad óptica del cultivo (que proporcionan una aproximación cuantitativa a la multiplicación de las bacterias dentro del cultivo), permiten obtener la tasa de cambio de la cantidad de fluorescencia por unidad de densidad óptica. A partir de estos valores se puede calcular una actividad promedio del promotor dentro del conjunto de la población. La fluorimetría de cultivos puede paralelizarse en placas multipocillo, lo que permite la caracterización rápida de la actividad de un promotor ante distintos estímulos o condiciones de crecimiento. Por estas razones nuestro grupo planteó, hace unos años, un método general que hacía uso de esta técnica experimental para determinar el efecto cuantitativo de un regulador transcripcional sobre un promotor diana (Fernandez-Lopez *et al.*, 2010). Nuestro método se basaba en la expresión controlada del regulador (el *input*), cuya producción podía ser inducida añadiendo distintas concentraciones de arabinosa (Figura 1-6.B). Para cada uno de estos niveles de expresión, el valor del *output* se determinaba mediante fluorimetría de cultivos. En su descripción original, este método se empleó para parametrizar el efecto del regulador transcripcional KorA, lo que nos llevó a proponer que este método podría ser adecuado para la parametrización sistemática de las redes de regulación transcripcional en bacterias. Por ejemplo, este tipo de análisis permitió determinar la red de regulación del plásmido R388 (Fernandez-Lopez *et al.*, 2014).

Durante el desarrollo de esta tesis, la aplicación de este método al par formado por el regulador TetR y su promotor diana p_{tetA} produjo, sin embargo, resultados desiguales. Por un lado, la actividad transcripcional desregulada de los promotores pudo determinarse correctamente (Figura 3-1). El análisis de la actividad junto con el promotor de referencia BBa_J23101, propuesto como estándar de actividad *in vivo* (Kelly *et al.*, 2009), permite relativizar las fuerzas de los distintos promotores y contextualizarlas para poder comparar entre laboratorios y técnicas.

Respecto al mecanismo de regulación, la técnica fue capaz de discernir entre el estado libre y reprimido del promotor p_{tetA} . Sin embargo, el registro de la dinámica de la represión se vio limitado por la sensibilidad. Para el caso del vector de expresión pBAD33, la cantidad de regulador producido, incluso en condiciones represoras para el sistema de expresión, bajaba la actividad de p_{tetA} a niveles inferiores a los detectables por el fluorímetro. Como solución, el cambio del sistema de expresión a un vector con un número de copias menor supuso una expresión basal de TetR menor, y se pudieron medir gradientes de actividad transcripcional en diferentes condiciones de cultivo (Figura 3-5).

No obstante, el cálculo de la función de transferencia necesita, no solo de la expresión controlada del *input*, sino de su cuantificación. En (Fernandez-Lopez *et al.*, 2010) se utilizó una estimación aproximada de los niveles de regulador correspondiente a los distintos niveles de arabinosa. Para ello, se calibró la respuesta del promotor p_{BAD} a la arabinosa utilizando *gfp* como gen reportero. Para cada nivel de arabinosa se obtenía así un nivel de GFP, y la curva entre arabinosa/GFP producida se asumía representativa de la inducción de una proteína cualquiera bajo pBAD. En el caso de TetR, esta estimación fue imposible. Las inducciones por arabinosa que producían una respuesta de p_{tetA} detectable por fluorimetría no daban señal de actividad en la curva de calibración con el promotor p_{BAD} . Intentamos entonces medir directamente la concentración de regulador. Para ello realizamos una fusión traduccional del regulador a la proteína fluorescente roja mKate2. La fusión traduccional resultó en una proteína estable y funcional (Figura 3-6.A y B). Sin embargo, los niveles de fluorescencia producidos por mKate2 fueron insuficientes para su cuantificación (Figura 3-6.C). Cuando la proteína de fusión TetR:mKate2 alcanzaba niveles detectables, los niveles de GFP ya estaban totalmente reprimidos (Figura 3-7 y Figura 3-11). En resumen el rango de detección del *input* no solapaba con el rango de detección del *output* haciendo inviable el cálculo de la función de transferencia.

Datos no mostrados en esta tesis indican que el problema de la sensibilidad de la fluorimetría de cultivos puede ser frecuente. Por ejemplo, intentos para parametrizar los pares promotor/regulador cl/p_R del fago lambda y MerR/ p_{mer} de Tn21 resultaron infructuosos (ensayos llevados a cabo por nuestro grupo, datos sin publicar). En función de la fuerza del promotor y de la afinidad del regulador por su operador, los valores de ambas señales pueden quedar fuera del rango de sensibilidad de la técnica.

4.3. La microscopía de fluorescencia permite resolver el problema de la sensibilidad experimental

Dado que el problema fundamental para la determinación de la función de transferencia parecía estar en la sensibilidad de la fluorimetría de cultivos, decidimos analizar nuestros sistemas experimentales utilizando una herramienta más sensible: el microscopio de epifluorescencia. La microscopía de fluorescencia requirió importantes cambios en nuestro protocolo experimental. En primer lugar, en la fluorimetría la medida de los valores totales de fluorescencia y número de células permite obtener de manera inmediata el promedio poblacional (simplemente dividiendo $\text{GFP}/\text{OD}_{600}$). En cambio, la microscopía nos proporciona valores individuales. Para poder estimar los valores poblacionales de manera adecuada necesitamos un muestreo suficiente. Dado que en un campo de microscopía tan solo es posible seguir un número muy limitado de células a lo largo del tiempo, decidimos recurrir al método de “instantánea” (*snapshot*). Este método consiste en tomar células que se encuentran creciendo en cultivo líquido en fase exponencial, colocarlas sobre un portaobjetos que contiene un soporte de agarosa con las mismas concentraciones de nutrientes, e inmediatamente tomar una instantánea en el microscopio. Realizando múltiples instantáneas pudimos obtener un muestreo suficiente de la población, tomando imágenes de varias decenas de miles de células por sistema experimental analizado.

Otra diferencia significativa respecto a la fluorimetría es que la cuantificación de la señal en microscopía requiere de un calibrado mayor. En nuestro caso, nos aseguramos de corregir las aberraciones en la señal producidas por la curvatura de la lente (*flat field correction*) así como de calibrar exposiciones y niveles de intensidad (Figura 2-9, Materiales y Métodos). Para cada nivel de inducción se obtuvieron imágenes y se segmentaron entre 1200 y 2000 células.

El paso a la microscopía de epifluorescencia supuso una mejora en los umbrales de detección. En los tres sistemas genéticos analizados se pudo correlacionar la señal del *input* con la señal del *output* en un rango dinámico que abarca entre 2 y 4 logs (Figuras 3-16, 3-19 y 3-23).

4.4. El comportamiento de la media y la mediana frente a la dispersión poblacional

El microscopio no solo nos proporcionó la sensibilidad necesaria para medir la relación *input/output* en todo el rango de concentraciones, sino que además nos permitió comparar los

valores poblacionales con los estadísticos que describen el comportamiento global de la población, como la media y la mediana. En las Figuras 3-16, 3-19 y 3-23 podemos observar que, en los tres sistemas experimentales, la media se desvía drásticamente de la mediana en las regiones de transición entre los estados encendido y apagado del promotor. Esta diferencia, que llega a ser de hasta 13 veces para un mismo valor de *input*, refleja el tremendo impacto que la estocasticidad en la expresión genética tiene sobre la función de transferencia. En su análisis de la función de transferencia, Rosenfeld *et al.* ya advertían que las medidas poblacionales podrían dar una imagen sesgada de la función de transferencia (Rosenfeld *et al.*, 2005). Nuestros resultados indican que este sesgo puede ser realmente alto: en muchos tramos el promedio poblacional, arrastrado por los valores más altos de la distribución, se aparta de la masa de probabilidad y se sitúa en regiones del plano donde no hay prácticamente ninguna célula. La naturaleza de la mediana, menos sensible a los valores extremos, amortigua en cierto modo esta tendencia, por lo que esta debe ser considerada un estadístico más adecuado para describir el comportamiento de la población. En cualquier caso, nuestros datos indican que, de haber sido posibles, las medidas poblacionales obtenidas por fluorimetría estarían altamente condicionadas por los valores superiores de la distribución de *outputs*. Como apuntaban Rosenfeld *et al.* (Rosenfeld *et al.*, 2005), esto daría lugar a una infraestimación considerable de la linealidad de la respuesta. Nuestros resultados indican que la cooperatividad (n) es sistemáticamente mayor en los ajustes para la mediana que en los ajustes para la media. Para el sistema de dos plásmidos, n tiene un valor 3,69 para la media y 3,83 para la mediana. En el caso del sistema con un solo plásmido (pRRG63) n es 3,33 para la media y 3,59 para la mediana. Y en el sistema integrado en el cromosoma (RRS247) n es 1,44 para la media y 1,49 para la mediana. Dado que incluso la mediana aparece sesgada por los valores superiores, los valores de n ajustados para este estadístico deben tomarse como un límite inferior de la linealidad real de la respuesta.

4.5. Diferencias en el rango de la función de transferencia debidas a la localización genética del sistema

Nuestros resultados muestran las distribuciones de *input/output* cuando el par TetR / p_{tetA} se encuentra en tres configuraciones genómicas distintas. Por un lado, tenemos un sistema en el que un *cassette* genético que contiene tanto el regulador como el promotor regulado se encuentra clonado dentro del cromosoma de *E. coli*, en el locus del operón de la arabinosa (RRS247). Este mismo *cassette* también se insertó en un plásmido de bajo número de copias

(pRRG63) y en otra construcción se clonó en plásmidos distintos la región del *input* y la región correspondiente al *output* (pRRG62 y pRRG13). El análisis de los mapas de *input/output* para estos tres sistemas reveló diferencias significativas.

Por un lado, el rango de valores de *input* y *output* fue diferente para los tres sistemas, aun cuando se utilizaron niveles de inducción de TetR equivalentes. Definimos R como el rango de represión, equivalente a dividir los niveles medianos de *output* en condiciones de no represión frente a los niveles medianos de *output* cuando la represión es total (niveles de TetR saturantes). Comparando los niveles de R obtenidos para los tres sistemas experimentales, observamos que el sistema cromosómico mostró un nivel significativamente menor ($R=80,2$) frente al obtenido para el sistema de un plásmido ($R=537,7$) y al sistema de dos plásmidos ($R=379,2$). Esto es debido a que la localización plasmídica de ambos sistemas hace que el nivel desregulado (en ausencia de TetR) alcance niveles mayores. Dado que el nivel completamente reprimido fue similar en los tres casos, el resultado es una R menor para el sistema cromosómico, que tiene un menor número de copias del gen por célula. Los diferentes niveles de R obtenidos para el sistema de un plásmido (Figura 3-20) y el sistema de dos plásmidos (Figura 3-17) se deben, probablemente, al distinto número de copias de los plásmidos utilizados. En el sistema de dos plásmidos el *output* ($p_{tetA}:gfp$) se clonó en un vector con origen de replicación de SC101, que mantiene unas 5 copias de plásmido por cromosoma (Hasunuma and Sekiguchi, 1977). En el sistema de un plásmido (pRRG63) el origen de replicación RK2 presenta un número de entre 4 y 7 copias de plásmido por cromosoma (Figurski *et al.*, 1979). En nuestras condiciones de cultivo RK2 presenta un mayor número de copias que SC101, como demuestran los niveles de fluorescencia de los controles con pRRG13 (SC101) y pRRG85 (RK2) en la Figura 3-9. Como consecuencia, el sistema de dos plásmidos mostró un nivel mayor de TetR por promotor p_{tetA} , reduciendo el rango de expresión entre los valores desreprimidos y completamente reprimidos de p_{tetA} .

4.6. Efecto de la localización genética del sistema sobre las fluctuaciones

La localización del sistema en plásmidos o en el cromosoma no solo produjo diferencias en el rango de regulación, sino que tuvo también un impacto evidente en los niveles de fluctuación de la relación *input/output*. En su análisis de los factores que condicionan el nivel de fluctuación en la expresión genética, Paulsson apuntó a la localización plasmídica como una posible fuente de ruido (Paulsson, 2004). Los plásmidos son moléculas de DNA de replicación independiente y no sincronizada con la replicación cromosómica, que se encuentran en un

número variable dentro del citoplasma celular. La asincronía respecto a la replicación cromosómica hace que el número de copias de los plásmidos sea fluctuante (Lau *et al.*, 2013). Aunque los plásmidos contienen mecanismos destinados a limitar estas fluctuaciones (Lestas *et al.*, 2010), en general desconocemos su magnitud respecto al número de equivalentes cromosómicos. Este ruido en el número de copias se transmite a aquellos genes codificados por el plásmido (Chabot *et al.*, 2007) y se espera que tenga un impacto sobre la capacidad de los mecanismos de control de la célula (Raser and O'Shea, 2005). Nuestros resultados permiten comparar las fluctuaciones en la relación *input/output* entre TetR y su promotor regulado en tres condiciones distintas: cuando *input* y *output* se encuentran en el cromosoma (RRS247), cuando *input* y *output* se encuentran en un plásmido que experimenta fluctuaciones en su número de copias (pRRG63) y cuando *input* y *output* se encuentran en plásmidos distintos que experimentan fluctuaciones independientes (pRRG13+pRRG62). Al comparar los niveles de ruido (expresados como el coeficiente de variación en el *output*, GFP) frente a los niveles promedio de *input* (valores promedio de TetR), podemos observar que los tres sistemas dieron lugar a perfiles diferentes.

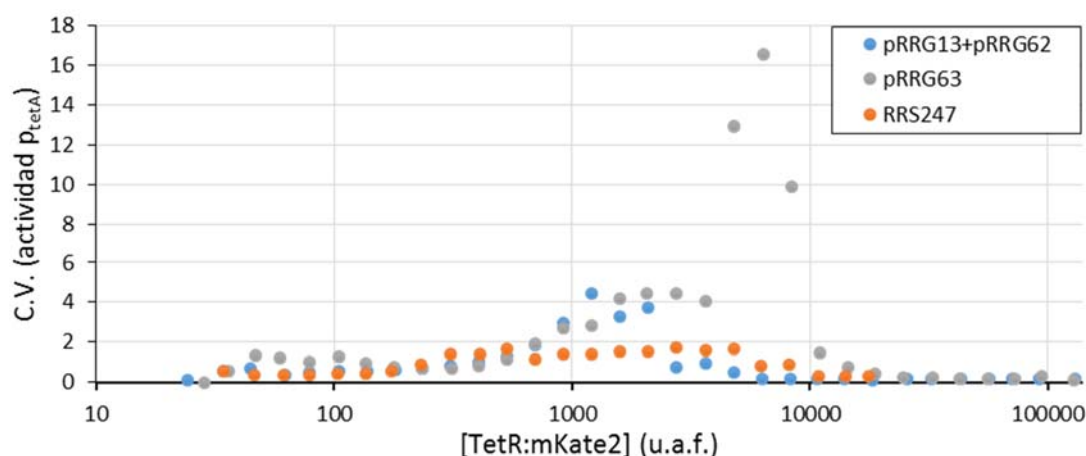


Figura 4-1. Coeficientes de variación de la actividad promotora de p_{tetA} en función de los rangos de concentración de TetR. Los datos representados pertenecen a la media de los *bins* de $[TetR:mKate2]$ en los experimentos de las Figuras 3.16 (pRRG13+62), 3.19 (pRRG63) y 3.23 (RRS247).

Por un lado, el sistema cromosómico (RRS247) mostró, globalmente, menores coeficientes de variación en el *output* para todo el dominio de concentraciones del *input*. Este resultado va en la línea de lo predicho por los análisis teóricos (Paulsson, 2004), indicando que la localización cromosómica se traduce en menores niveles de ruido. Por otro lado, pese a que en principio eran esperables mayores niveles de ruido en la respuesta cuando *input* y *output* fluctúan de manera independiente (pRRG13+pRRG62), los datos indican un panorama distinto. La comparación entre el sistema de uno y dos plásmidos indica que los niveles de ruido incrementan de manera similar en ambos sistemas al acercarse a la K aparente (ajustes de las

Figuras 3-17 y 3-20). Obviando los tres valores máximos para el sistema de un plásmido (ya que incorporan al valor propio de su dispersión un artefacto debido a una región inframuestreada), valores de TetR superiores a K produjeron mayores niveles de ruido en el sistema de un plásmido, a pesar de que en este sistema se espera que las fluctuaciones en el número de copias del gen de TetR y de p_{tetA} sean sincrónicas (dado que ambos genes están codificados en el mismo plásmido). No es fácil dar una hipótesis obvia a este fenómeno: la diferencia en el número de copias esperado entre los plásmidos utilizados es demasiado pequeña como para ser la única causa (en el caso de SC101 las fluctuaciones escalarían respecto a 1/5 frente a 1/7 en el caso de RK2). Es posible que ambos plásmidos tengan sistemas de control del número de copias diferentes, que hacen que para una media similar, RK2 presente fluctuaciones menores.

4.7. Ajustes de la función de transferencia a la mediana poblacional: ¿una o dos fases?

Uno de los resultados más sorprendentes en nuestros mapas de dispersión de *input/output* fue que, tanto en el caso del sistema cromosómico como en el sistema de un plásmido, la mediana de la concentración de GFP parecía seguir una respuesta bifásica frente a la concentración de TetR. En ambos casos la curva sugería dos eventos de represión concatenados. Una posibilidad es que esta respuesta bifásica correspondiese a la unión a cada uno de los dos operadores *tetO* presentes en el promotor de p_{tetA} . Como se detalla en la introducción los distintos estudios publicados, aunque caracterizan la afinidad de ambos operadores en el mismo orden de magnitud, llegan a asignar hasta 5 veces más afinidad a *tetO*₂ (Kleinschmidt *et al.*, 1991). En nuestros ajustes, la diferencia entre la afinidad de los dos posibles eventos de represión fue: entre 3-4 veces para el vector pRRG63; y entre 13-14 veces para la cepa RRS247 (cociente entre las K s de los dos eventos, K_2/K_1 , ajustadas en las Figuras 3-21 y 3-25). En el caso del sistema de dos plásmidos, no se apreció ningún comportamiento bifásico en la mediana. Esto resta verosimilitud a esta explicación mecánica: independientemente del número de copias y de la localización cromosómica, el mecanismo de regulación de TetR debería ser el mismo. Por otra parte, las simulaciones apuntan a una explicación más probable.

Utilizando las simulaciones basadas en CMEs, tal y como se describen en Materiales y Métodos, podemos determinar hasta qué punto el ajuste de la mediana puede estar condicionado por las fluctuaciones del sistema. Para ello, simulamos el comportamiento de un sistema represor/gen regulado con $n=4$ y $K=100$ unidades. Sometimos al sistema a 10

concentraciones logarítmicamente crecientes de regulador, y obtuvimos los valores de *input/output* para las proteínas del sistema (X, Y). A partir de esta nube, realizamos un tratamiento similar al de los datos experimentales, intentando recuperar los valores originales de K y n que introdujimos en la simulación. Como puede observarse en la Figura 4-2, la media móvil a lo largo del eje X revela un comportamiento polifásico, en lugar del comportamiento esperado para la función que introdujimos en la simulación. Ajustando el valor de la mediana a una única función, obtuvimos una K de 87 y una n de 3.4, valores inferiores a los valores reales ($K=100, n=4$) que introdujimos en la simulación. Este resultado refuerza la idea de que incluso la mediana es un estadístico que infraestima la linealidad de la respuesta. Tal y como sucede para los resultados experimentales, es posible, además, ajustar la mediana a una curva bifásica obteniendo altos valores de R^2 (0.996 y 0.9639), pero sin representar en absoluto el mecanismo real subyacente, dando unas K s de 115 y 382 con n s de 6 y 5 respectivamente.

El mapa de dispersión X, Y nos revela el origen de este comportamiento anómalo de la mediana. Cuando inducimos el regulador obtenemos distintos valores promedio, y amplias zonas del rango de X se pueblan gracias a la dispersión en X , consecuencia del ruido en la expresión. Dependiendo de la magnitud del ruido y la distancia entre los valores promedio de X que inducimos (en nuestros experimentos, debidos a los diferentes niveles de arabinosa utilizados), obtenemos zonas con baja densidad de muestreo. En estas zonas, la mediana se encuentra totalmente dominada por los valores superiores, lo que da lugar a saltos que pueden producir una aparente curva polifásica.

Tal y como se detalla en Resultados, las fluctuaciones observadas en nuestras simulaciones deben tomarse como un límite inferior a la situación real, puesto que *in silico* no introdujimos fluctuaciones extrínsecas (debidas a las variaciones en RNA polimerasas, ribosomas, etc.), que en la célula son probablemente muy importantes a la hora de condicionar el ruido en la expresión genética (Eldar and Elowitz, 2010; Raser and O'Shea, 2005). No obstante, las simulaciones indican que no es necesario invocar mecanismos anómalos de regulación: la apariencia bifásica de la mediana en los sistemas cromosómico y de un plásmido puede explicarse debido a la existencia de zonas de TetR inframuestreadas.

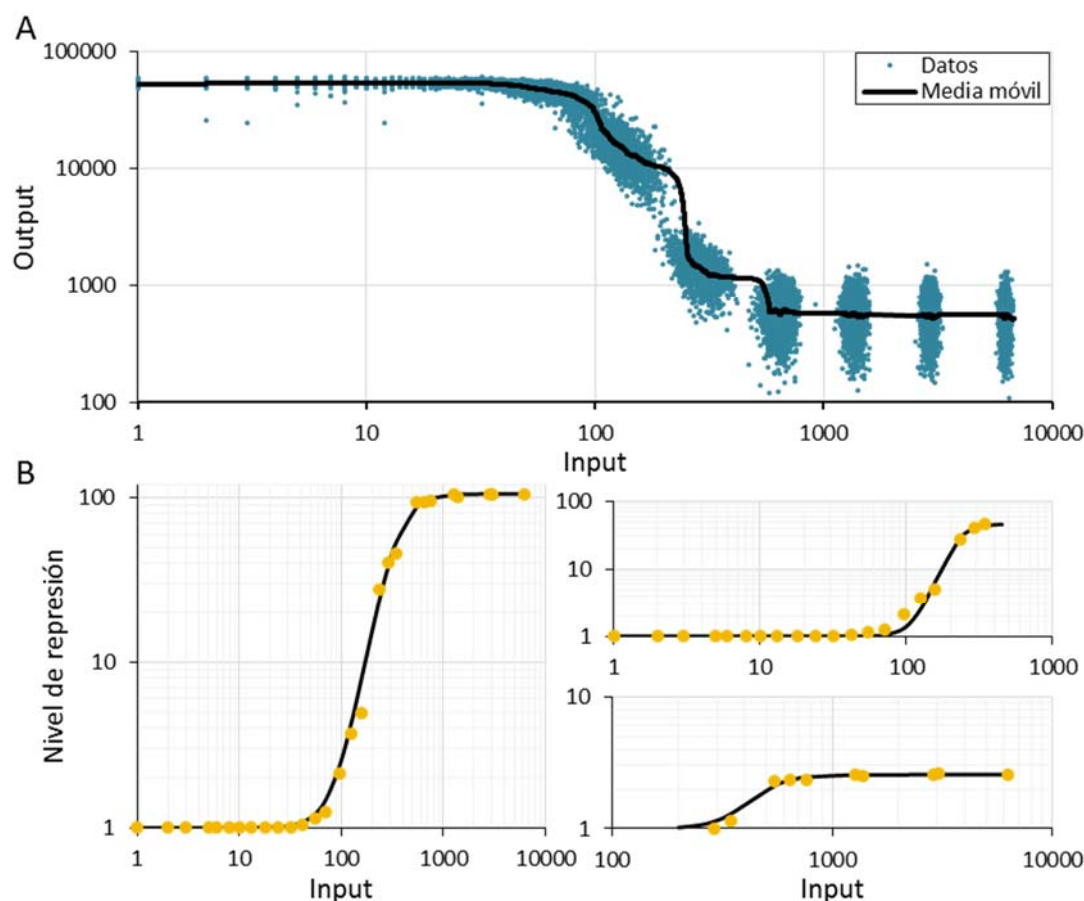


Figura 4-2. Simulación utilizando el código de Matlab (apartado 2.9.2) de un muestreo con diferentes condiciones de inducción. (A) Representación del *output* (en nuestro sistema, actividad promotora) frente al *input* (en nuestro sistema, concentración de regulador). La media móvil (línea negra) muestra un desarrollo multifásico consecuencia de las diferentes poblaciones (nubes de puntos) producidas por el método de muestreo. (B) Representación del nivel de represión del *output* (Y_{max}/Y_i) en función del *input*. Ajuste de la mediana a la Ecuación 3-1 (función de transferencia) en uno y dos rangos del *input*.

4.8. Divergencia entre los datos bioquímicos y los resultados in vivo

Las simulaciones indicaron que el ajuste más probable para la función de transferencia entre TetR y su promotor diana p_{tetA} corresponde a una única constante y una cooperatividad aparente en torno a 4. Los resultados bioquímicos, obtenidos *in vitro*, sin embargo, apuntan a una n diferente. La afinidad de los dímeros de TetR a los operadores ha sido descrita como no cooperativa (Kleinschmidt *et al.*, 1991; Kleinschmidt *et al.*, 1988), por lo que, suponiendo que la dimerización de TetR ocurriese en concentraciones similares a las de la unión a sus operadores, esperaríamos una cooperatividad aparente de $n=2$. Incluso este límite es difícilmente justificable de acuerdo a los datos bioquímicos: la constante de dimerización de TetR está 5 órdenes de magnitud por debajo, de acuerdo a la literatura (Kleinschmidt *et al.*, 1991; Kleinschmidt *et al.*, 1988).

De todas formas, no hay que olvidar que en nuestros resultados n es tan solo una medida de la linealidad de la respuesta, por lo que la discrepancia entre los resultados *in vivo* y los resultados *in vitro* lo que nos indican es que el proceso de represión dentro de la célula es más complicado y menos lineal. Una posible causa para esta discrepancia radica en las diferentes condiciones de unión que la proteína experimenta *in vivo* o *in vitro*. En el tubo de ensayo las concentraciones de proteína y operador son constantes, no fluctuantes y además sometidas a difusión en 3 dimensiones. Por otra parte dentro de la célula los reguladores deben hacer frente al *crowding* molecular, experimentan difusión en 1 dimensión a lo largo del DNA y están sometidos a altos niveles de fluctuación en sus concentraciones. Además, mientras que los ensayos bioquímicos se realizan en condiciones de equilibrio, el equilibrio es solo una asunción en los ensayos *in vivo*. Estudios realizados mediante microscopía estroboscópica, capaces de medir *in vivo* la cinética de unión de un TF, apuntan a que la condición de equilibrio rápido rara vez se cumple, siendo la constante de disociación el factor dominante en la cinética (Elf *et al.*, 2007).

4.9. Control analógico frente a control digital

Una de las cuestiones fundamentales que aún permanecen abiertas en torno a la regulación transcripcional, es la capacidad de la función de transferencia de transmitir información respecto a las concentraciones de un regulador dado hasta los genes regulados por este. Como indicamos en la introducción, existen dos límites teóricos respecto a la sensibilidad de un *output* respecto a las variaciones en el *input*. Por un lado, existe la posibilidad de que variaciones en la concentración del regulador produzcan inexorablemente variaciones en la regulación del gen regulado. Cuando cualquier variación en el *input* (dx) produce una variación del *output* (dy) para cualquier valor del primero (x), nos encontramos en un límite que denominaremos control analógico. Un controlador analógico es capaz, por tanto, de reportar información sobre la magnitud de la señal. Por el contrario, podría suceder que el gen regulado solo respondiese a aquellas variaciones en el *input* que superasen un determinado valor umbral. A este tipo de regulación, por analogía con los circuitos electrónicos, lo denominaremos control digital. La existencia de fluctuaciones estocásticas en la expresión de los genes añade, además, un nivel adicional de complejidad. En un entorno sometido a cierta estocasticidad, un controlador digital puede mostrar heterogeneidad a la hora de alternar entre sus estados encendido/apagado. Esto es, ante unas mismas condiciones, las concentraciones de regulador estarán distribuidas de una determinada manera. Si la dispersión en esta distribución es lo suficientemente amplia, unas células podrían encontrarse en un lado u otro del umbral de activación. Cambios en la distribución de regulador

producirían entonces un cambio en la probabilidad de activación de cada célula individual. Como se muestra en la Figura 4-3, a nivel poblacional no es posible diferenciar entre un controlador analógico y un controlador digital sometido a ruido. Dado que el comportamiento poblacional es la resultante de la suma de los comportamientos individuales, el nivel total observado mostrará cambios en la magnitud promedio del *output* ante cualquier cambio en la intensidad promedio del *input*. Esta equivalencia a nivel macroscópico no quiere decir, sin embargo, que ambos mecanismos de control sean equivalentes a nivel microscópico: en el caso de un controlador analógico las células individuales son capaces de cuantificar los cambios en la señal. En el caso de un controlador digital, mientras que la población en conjunto muestra respuestas cuantitativas a los cambios en la magnitud de la señal, las células individuales tan solo son capaces de distinguir entre dos estados posibles.

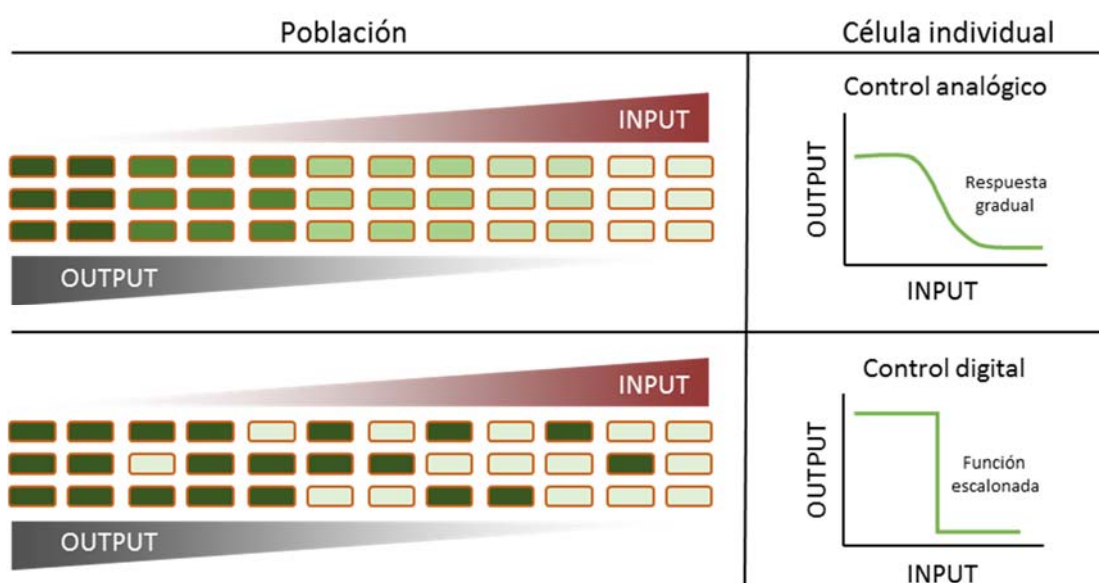


Figura 4-3. Esquema del comportamiento poblacional frente al individual. El comportamiento gradual que presenta la población puede deberse tanto a un control digital como analógico a nivel celular. En la parte izquierda de la figura se representan las células de la población en distintos grados de producción del *OUTPUT* (distintos tonos de verde) en respuesta al *INPUT*. A la derecha de la figura se representa la función de respuesta (*OUTPUT* frente a *INPUT*) para una célula individual que se comporta con un control analógico (parte superior) o digital (parte inferior).

Nuestros resultados experimentales para TetR mostraron que tanto la media como la mediana poblacional mostraban cambios proporcionales a la magnitud de la concentración de TetR. Sin embargo, también indicaron que media y mediana presentaban problemas a la hora de representar el comportamiento típico de las células individuales. En amplios rangos de concentración de TetR, las fluctuaciones en torno a la mediana fueron de varios órdenes de magnitud (Figura 3-19). En estas condiciones no está claro que el ruido en la expresión genética permita una discriminación no ambigua a nivel individual de las concentraciones de regulador. Para poder contestar a esta cuestión necesitamos cuantificar de alguna manera la

incertidumbre en la relación *input/output*, esto es, para determinar el modo de control de TetR, necesitamos determinar cuánta información transmite la concentración de TF a sus genes regulados.

4.10. Información mutua en la regulación por TetR

Bialek y colaboradores (Gregor *et al.*, 2007a) fueron los primeros en proponer que la incertidumbre en la relación *input/output* para un TF y su gen regulado no era otra cosa que la información mutua, tal y como fue definida por Shannon. Siguiendo la definición de Shannon, la medida de la información mutua entre un TF y su gen regulado requiere medir las entropías en las distribuciones marginales de *input* $P(X)$, *output* $P(Y)$, así como la entropía de la distribución conjunta $P(X, Y)$. La reducción en entropía de la distribución conjunta respecto a la suma de las entropías marginales es, entonces, una medida de la información mutua entre X e Y :

$$I(X; Y) = H(X) + H(Y) - H(X, Y)$$

En nuestro caso, esto requería medir la respuesta de p_{tetA} , expresada en unidades de GFP, ante todas las concentraciones de TetR que la célula fuese capaz de experimentar. Esta última condición plantea un problema importante: aunque nuestro sistema de expresión permite saturar la célula de TetR, concentraciones crecientes de TetR por encima del rango fisiológico producirían una disminución artefactual de la información mutua. Un aumento de TetR por encima de su rango fisiológico produciría un aumento de la entropía de X (que es una medida del número de concentraciones posibles) mientras que $H(X, Y)$ permanecería constante, dado que a partir de un valor saturante el regulador no disminuye la producción de gen regulado. Para resolver este problema, utilizamos los valores obtenidos cuando TetR se encuentra bajo su promotor nativo, p_{tetR} . Dado que en el sistema TetR el inductor actúa rebajando las concentraciones de regulador capaz de unirse a su operador, la concentración máxima de TetR experimentada por la célula debe ser, necesariamente, aquella producida por su propio promotor. De esta manera, pudimos poner un límite superior a $H(X)$.

Un segundo problema a la hora de calcular la información mutua, tal y como se describe en Materiales y Métodos, reside en la transformación de las nubes de dispersión en masas de probabilidad. Para poder asignar probabilidades a partir de datos individuales necesitamos dividir el espacio (X, Y) en distintas categorías o *bins*. Bialek y colaboradores ya establecieron que asignar *bins* de manera fisiológica resulta problemático (Gregor *et al.*, 2007a). Siguiendo sus pasos, nuestra estrategia consistió en calcular la información mutua para un número

creciente de *bins*, tomando el máximo obtenido como el límite superior para la *I*. De esta forma, pudimos obtener los valores máximos de información entre TetR y p_{tetA} cuando el sistema se encontraba insertado en el cromosoma (Figura 3-27.C), en un plásmido (Figura 3-27.B) y dividida en dos plásmidos diferentes para *input* y *output* (Figura 3-27.A).

4.11. TetR actúa como un controlador digital ruidoso

Aplicando esta metodología obtuvimos valores de $I(TetR; GFP)$ de alrededor de 1 bit, tanto para el sistema cromosómico, como para los sistemas de uno y dos plásmidos. Dado que el número de bits nos indica el número de niveles que el sistema es capaz de distinguir sin error, este valor hace de TetR un controlador digital ruidoso. La célula, por tanto, es capaz de diferenciar dos niveles de TetR, aunque bajo cierto error de inferencia. Las distribuciones de probabilidad (Figura 3-27, paneles de la izquierda) nos muestran dos regiones correspondientes a un nivel activado y un nivel desactivado de p_{tetA} , unidas por una región en torno a la *K* aparente en el que las fluctuaciones hacen imposible la asignación exacta del estado de activación: hay células que muestran niveles equivalentes al estado activado, otras al estado inactivado y un tercer grupo que muestra valores intermedios. Resulta interesante resaltar que, si bien las diferentes localizaciones genómicas del sistema produjeron diferencias en el nivel de ruido en el *output* (Figura 4-1), este aumento en las fluctuaciones no se tradujo en una disminución apreciable de la información mutua. Como apuntamos anteriormente, las distintas localizaciones genómicas produjeron también un cambio en el rango de expresión del gen regulado. Un aumento en el rango de expresión del gen regulado llevaría acompañado un aumento en la información mutua (Bowsher and Swain, 2014; Dubuis *et al.*, 2013; Hansen and O'Shea, 2015; Tkacik *et al.*, 2009), por lo que es probable que el aumento en el rango del *output* compense la disminución en la capacidad informativa producida por un aumento del ruido en la expresión.

En cualquier caso, los valores de *I* obtenidos indican que la célula tan solo es capaz de identificar (y con cierto error asociado) la presencia o ausencia de TetR por encima o debajo de un umbral. Las distribuciones representadas en las Figuras 3-17, 3-20, 3-24 indican que este valor umbral se sitúa alrededor de *K*, pero las fluctuaciones del sistema hacen que sea más adecuado hablar de una ventana de transición que de un nivel *sensu stricto*. Valores de TetR situados en esta ventana producen *outputs* en p_{tetA} a lo largo de toda su distribución posible, con células que para un mismo nivel de TetR muestran diferencias de casi 3 órdenes de magnitud en el *output*.

4.12. La desigualdad en el procesamiento de la información

Una vez que obtuvimos los valores de información mutua de manera experimental, nos preguntamos si nuestras simulaciones podrían arrojar luz sobre los factores que condicionan la capacidad informativa de la función de transferencia. El resultado más sorprendente de nuestro análisis computacional fue la inesperada violación de la desigualdad en el procesamiento de la información que observamos a lo largo del canal transcripcional. La DPI establece que, dado una serie de procesos markovianos, la información a lo largo de un canal de información puede disminuir o mantenerse, pero ninguna acción local puede aumentar la información. En nuestro caso concreto, esto significaría que, dado que TetR actúa regulando la síntesis de mRNA a partir de p_{tetA} , y en nuestro sistema experimental GFP se traduce a partir de este mRNA, la información entre TetR y el mRNA producido a partir de p_{tetA} debería ser igual o mayor que la información entre TetR y GFP. Nuestro sistema experimental no permite medir los niveles de mRNA, pero nuestras simulaciones sí permiten calcular su distribución. Es importante resaltar, que, dado que en las simulaciones la hipótesis markoviana es una condición *ex ante*, en teoría $I(TetR; mRNA_{gfp})$ debería ser mayor o igual a $I(TetR; GFP)$. Como se muestra en la Figura 3-31 nuestros resultados indicaron lo contrario. Esto es, incluso aunque los procesos de transcripción/traducción fuesen perfectamente markovianos, asunción ampliamente establecida en la literatura (Lestas *et al.*, 2010; Paulsson, 2004; Swain *et al.*, 2002), un circuito transcripcional puede violar la DPI.

La razón para esta aparente incongruencia es que el regulador transcripcional actúa sobre la tasa de síntesis de mRNA, pero incluso aunque este proceso sea perfectamente markoviano, la concentración de mRNA y proteínas depende tanto de la síntesis como de la degradación. La suma de dos procesos markovianos produce un tercer proceso markoviano solo si ambos son procesos independientes. Dado que la degradación depende de las concentraciones, y estas dependen a su vez de las tasas de síntesis, síntesis y degradación no son procesos mutuamente independientes. De hecho, en una serie de procesos de Markov tales que

$$x \rightarrow y \rightarrow z$$

La DPI requiere que $X \perp Z|Y$. Esto es, que dado un valor de Y , X sea ortogonal a Z . En nuestras simulaciones (y con toda probabilidad, en la fisiología celular) esta condición no se cumple debido al diferente orden de magnitud en las tasas de degradación de mRNA y proteínas. Mientras que, en bacterias, el mRNA tiene vidas medias cortas del orden de minutos, las

proteínas pueden tener vidas medias del orden de horas (o días, en el caso de proteínas fluorescentes). Esto hace que las concentraciones de proteína muestren entre sí correlaciones de más largo alcance que las correlaciones entre mRNA y proteína, aun cuando mRNA y proteína pertenezcan a un mismo gen. Imaginemos por ejemplo qué sucedería si bloqueásemos completamente la síntesis de TetR en un tiempo t determinado. Suponiendo que para t los valores de TetR y GFP correspondiesen al valor en estado estacionario, existe una correlación temporal entre ambos valores debido al hecho de que ambos desaparecen a una misma velocidad. Por el contrario, la concentración de mRNAs sufre unas variaciones mucho más rápidas. Esta diferencia de “memoria” puede dar lugar a una mayor información mutua entre proteínas, aun cuando el nexos causal sea la producción de mRNA.

4.13. Efecto de la velocidad de degradación en la información mutua

Esta dependencia de la información mutua de la vida media de los componentes del circuito nos plantea una duda. Nuestro sistema experimental se basa en fusiones traduccionales o transcripcionales a proteínas fluorescentes (GFP, mKate2) con una larguísima vida media, que puede llegar a los varios días (Andersen *et al.*, 1998). Es posible que los componentes del sistema natural no tengan niveles semejantes de estabilidad, lo que podría afectar a la información mutua real entre TetR y su gen regulado *tetA*. Para analizar esta posibilidad realizamos simulaciones en las que modificamos sistemáticamente la vida media de las proteínas del sistema. Los resultados (Figura 3-34) indicaron que el máximo de información mutua se consigue con proteínas de larga vida media. Dado que las proteínas fluorescentes son esencialmente estables, la vida media de los componentes en el sistema natural puede ser igual o menor que la de nuestras fusiones fluorescentes. Esto hace que la información mutua obtenida experimentalmente deba considerarse el límite superior de I para el sistema regulador de TetR.

4.14. La información transmitida por TetR frente a otros reguladores transcripcionales

Nuestros resultados representan la primera medida efectiva de la información mutua para un regulador transcripcional procariótico, por lo que resulta difícil establecer una comparación entre la capacidad informativa de TetR frente a otros reguladores transcripcionales similares. En organismos eucarióticos, los únicos casos en los que la I ha podido ser medida de manera cuantitativa se circunscriben al activador transcripcional Hunchback, que regula la génesis del

eje anteroposterior del embrión de la mosca del vinagre (Dubuis *et al.*, 2013), así como a los reguladores humanos ERK y NF- κ B (Selimkhanov *et al.*, 2014) y al regulador MS2 de levaduras (Hansen and O'Shea, 2015). En los casos de ERK, NF- κ B y MS2, los resultados obtenidos indicaron informaciones mutuas cercanas, aunque inferiores a 1 bit (Hansen and O'Shea, 2015; Selimkhanov *et al.*, 2014), similares a nuestros resultados con TetR. Hunchback, sin embargo, dio valores de I alrededor de 1,8 bits, indicando que en el caso del eje anteroposterior de la larva de *Drosophila* existe cierta capacidad de inferir la concentración de regulador, lo que los autores correlacionaron con la necesidad de obtener información posicional (Dubuis *et al.*, 2013).

La similitud en los resultados para TetR, ERK, NF- κ B y MS2, factores de transcripción implicados en procesos muy diversos y procedentes de organismos radicalmente distintos, sugieren que el modelo del “controlador ruidoso” que hemos discutido para el caso de TetR podría resultar una forma de control ampliamente extendida entre distintos organismos. A este respecto hay que apuntar, no obstante, que existe cierta controversia acerca de la significación de los valores de I obtenidos para los reguladores eucarióticos, debido al peculiar modo de actuación de estos. Mientras que en bacterias regulador transcripcional y operador comparten un mismo espacio, los organismos eucariotas presentan una sofisticación adicional. En estos organismos, el factor transcripcional se encuentra frecuentemente en el citoplasma, y la unión al inductor produce su translocación al núcleo. Este proceso de translocación hace que los organismos eucarióticos presenten dos tipos de regulación transcripcional: modulada por amplitud y modulada por frecuencia. En los reguladores modulados por amplitud, el efecto del TF es similar al modelo procariótico, siendo la concentración total de regulador en el núcleo lo que condiciona la respuesta. En el modelo de regulación por modulación de frecuencia, la célula no responde a la concentración total de regulador sino a la frecuencia con la que pulsos de regulador son translocados al núcleo (Cai *et al.*, 2008). Un mismo regulador puede presentar promotores regulados por amplitud y otros regulados por frecuencia, como es el caso del factor MS2 en levaduras (Cai *et al.*, 2008; Hansen and O'Shea, 2015), o la proteína p53 en humanos (Purvis *et al.*, 2012). Esto hace que haya quienes postulen que las células recaban información a partir de la serie temporal, dado que el vector de concentraciones ofrece mayores niveles de I que los valores en estado estacionario (Selimkhanov *et al.*, 2014). Frente a esta opinión, Hansen y O'Shea sostienen que el incremento en información mutua es marginal, y que no está claro que la observación de un mayor nivel de información mutua aparente en dos distribuciones cualesquiera signifique

necesariamente que esas distribuciones sean, precisamente, las empleadas por la célula para la toma de decisiones (Hansen and O'Shea, 2015).

4.15. Optimización de la transmisión de información

En el fondo de este debate reside la cuestión de si las células optimizan de algún modo sus circuitos transcripcionales para transmitir el máximo de información posible. Bialek y Tkačik, en sus estudios sobre Hunchback concluyeron que dados los niveles de ruido en el sistema, la transmisión de información por Hunchback estaba cercana a su óptimo teórico (Dubuis *et al.*, 2013). Sin embargo, es difícil no encontrar cierta circularidad en este argumento, puesto que, como los propios autores determinaron, el factor fundamental en la capacidad informativa de Hunchback son las propias fluctuaciones en su concentración. Por otra parte, nuestros resultados y simulaciones indican que un nivel menor de cooperatividad podría mejorar la capacidad informativa de TetR, hecho que los resultados teóricos de Bialek y Tkačik para canales gaussianos también indicarían (Tkacik *et al.*, 2009).

En el fondo resulta difícil justificar que la evolución actúe optimizando inexorablemente los niveles informativos. Por un lado, la información no es un parámetro físico sobre el que la selección natural pueda actuar. Por otro lado, dado que las células no pueden discriminar qué molécula de regulador corresponde a señal y cual corresponde a ruido (Lestas *et al.*, 2010), un aumento en la sensibilidad a la señal implica indefectiblemente un aumento a la sensibilidad al ruido. Una posibilidad para evitar esta situación consistiría en la utilización de canales de información que operasen con alto número de moléculas, de forma que se minimizase el impacto de las fluctuaciones. Sin embargo, como apuntan Swain y colaboradores (Perkins and Swain, 2009), la célula debe tomar la decisión correcta en un ambiente fluctuante, frente a competidores que también toman decisiones y teniendo en cuenta la economía celular. Funcionar con alto número de moléculas implicaría un coste metabólico que podría superar la ventaja adaptativa de un sensor libre de fluctuaciones. En esta dirección, Leibler ha demostrado que, dependiendo del coste de la decisión y de la frecuencia con la que las células encuentran la señal, controladores estocásticos (incluso completamente aleatorios) pueden proporcionar valores superiores de *fitness* cuando la señal es suficientemente infrecuente (Rivoire and Leibler, 2011).

En el caso particular de TetR, esto podría indicar que la mejor estrategia para la célula en términos de *fitness* no corresponde a una optimización de la información mutua, sino a un controlador capaz de transmitir fielmente los niveles de antibiótico ante los cuales es más efectivo responder que no hacerlo. Para comprobar esta posibilidad, experimentos futuros

deberán determinar el rango de concentraciones de tetraciclina para los cuales la expresión de TetA resulta más deletérea que provechosa, y la correspondencia entre estos niveles y la capacidad informativa de TetR.

Conclusiones

Conclusiones

- Sobre la metodología para medir la función de transferencia

I. La microscopia de epifluorescencia proporcionó la sensibilidad necesaria para la cuantificación de la actividad del promotor p_{tetA} y las concentraciones de regulador TetR, que permitieron la determinación de su función de transferencia. Además, esta técnica permitió obtener las distribuciones poblacionales, necesarias para a) evaluar la dispersión poblacional y b) calcular la información mutua.

- Sobre la naturaleza de la función de transferencia

II. La función de transferencia para el par TetR/ p_{tetA} se caracterizó por una respuesta altamente sensible a la concentración de TetR (K de 0,014 veces el rango máximo de inducción) y un alto grado de no-linealidad ($n=4$). La dispersión poblacional en torno a la mediana fue muy alta, por lo que la función de transferencia requirió un segundo estimador que cuantificara el descenso en la incertidumbre en p_{tetA} en función de la concentración de TetR: la información mutua (I).

- Sobre la información mutua

III. La $I(TetR; GFP)$ de 1 bit indicó que el sistema de regulación por TetR funciona esencialmente como un controlador digital ruidoso. Junto a la trayectoria de la función de transferencia, este resultado indica que la célula es capaz de identificar la señal de TetR, pero no de cuantificarla.

IV. Tanto la forma de la función de transferencia como su información mutua asociada son valores necesarios para caracterizar la respuesta de un promotor a su regulador transcripcional. Estos parámetros deben ser determinados experimentalmente *in vivo* dado que: a) los valores de K y n obtenidos *in vitro* no se corresponden con los valores aparentes *in vivo* y b) la información mutua observada no corresponde, necesariamente, al óptimo informativo para unos niveles de ruido dados. Sensibilidad, resistencia al ruido y economía celular son factores que influyen en la evolución de la maquinaria celular responsable de la toma de decisiones.

Bibliografía

Bibliografía

- Andersen, J.B., Sternberg, C., Poulsen, L.K., Bjorn, S.P., Givskov, M., and Molin, S. (1998). New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl Environ Microbiol* 64, 2240-2246.
- Bell, C.E., and Lewis, M. (2000). A closer view of the conformation of the Lac repressor bound to operator. *Nat Struct Biol* 7, 209-214.
- Berens, C., and Hillen, W. (2004). Gene regulation by tetracyclines. *Genetic engineering* 26, 255-277.
- Bertrand, K.P., Postle, K., Wray, L.V., Jr., and Reznikoff, W.S. (1983). Overlapping divergent promoters control expression of Tn10 tetracycline resistance. *Gene* 23, 149-156.
- Bialek, W. (2012). *Biophysics: Searching for Principles* (Princeton University Press).
- Biliouris, K., Daoutidis, P., and Kaznessis, Y.N. (2011). Stochastic simulations of the tetracycline operon. *Bmc Syst Biol* 5.
- Bowsher, C.G., and Swain, P.S. (2014). Environmental sensing, information transfer, and cellular decision-making. *Curr Opin Biotechnol* 28, 149-155.
- Cai, L., Dalal, C.K., and Elowitz, M.B. (2008). Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* 455, 485-490.
- Carbonell-Ballesteros, M., Duran-Nebreda, S., Montanez, R., Sole, R., Macia, J., and Rodriguez-Caso, C. (2014). A bottom-up characterization of transfer functions for synthetic biology designs: lessons from enzymology. *Nucleic Acids Res* 42, 14060-14069.
- Cuthbertson, L., and Nodwell, J.R. (2013). The TetR family of regulators. *Microbiol Mol Biol Rev* 77, 440-475.
- Chabot, J.R., Pedraza, J.M., Luitel, P., and van Oudenaarden, A. (2007). Stochastic gene expression out-of-steady-state in the cyanobacterial circadian clock. *Nature* 450, 1249-1252.
- Chang, A.C., and Cohen, S.N. (1978). Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *Journal of bacteriology* 134, 1141-1156.
- Chen, X., Zaro, J.L., and Shen, W.C. (2012). Fusion protein linkers: property, design and functionality. *Advanced drug delivery reviews* 65, 1357-1369.
- Cherney, L.T., Cherney, M.M., Garen, C.R., and James, M.N. (2010). Crystal structure of the intermediate complex of the arginine repressor from *Mycobacterium tuberculosis* bound with its DNA operator reveals detailed mechanism of arginine repression. *J Mol Biol* 399, 240-254.
- Dangkulwanich, M., Ishibashi, T., Bintu, L., and Bustamante, C. (2014). Molecular mechanisms of transcription through single-molecule experiments. *Chem Rev* 114, 3203-3223.
- Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America* 97, 6640-6645.

- Dower, W.J., Miller, J.F., and Ragsdale, C.W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res* 16, 6127-6145.
- Dubuis, J.O., Tkacik, G., Wieschaus, E.F., Gregor, T., and Bialek, W. (2013). Positional information, in bits. *Proceedings of the National Academy of Sciences of the United States of America* 110, 16301-16308.
- Eckert, B., and Beck, C.F. (1989). Overproduction of transposon Tn10-encoded tetracycline resistance protein results in cell death and loss of membrane potential. *Journal of bacteriology* 171, 3557-3559.
- Eldar, A., and Elowitz, M.B. (2010). Functional roles for noise in genetic circuits. *Nature* 467, 167-173.
- Elf, J., Li, G.W., and Xie, X.S. (2007). Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* 316, 1191-1194.
- Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. *Science* 297, 1183-1186.
- Epe, B., and Woolley, P. (1984). The binding of 6-demethylchlortetracycline to 70S, 50S and 30S ribosomal particles: a quantitative study by fluorescence anisotropy. *Embo J* 3, 121-126.
- Fernandez-Lopez, R., Del Campo, I., Revilla, C., Cuevas, A., and de la Cruz, F. (2014). Negative feedback and transcriptional overshooting in a regulatory network for horizontal gene transfer. *PLoS genetics* 10, e1004171.
- Fernandez-Lopez, R., Del Campo, I., Ruiz, R., Lanza, V., Vielva, L., and de la Cruz, F. (2010). Numbers on the edges: a simplified and scalable method for quantifying the gene regulation function. *Bioessays* 32, 346-355.
- Fernandez-Lopez, R., Ruiz, R., de la Cruz, F., and Moncalian, G. (2015). Transcription factor-based biosensors enlightened by the analyte. *Front Microbiol* 6, 648.
- Figurski, D.H., Meyer, R.J., and Helinski, D.R. (1979). Suppression of Co1E1 replication properties by the Inc P-1 plasmid RK2 in hybrid plasmids constructed in vitro. *J Mol Biol* 133, 295-318.
- Gibson, D.G. (2011). Enzymatic Assembly of Overlapping DNA Fragments. *Method Enzymol* 498, 349-361.
- Gillespie, D.T. (1977). Exact stochastic simulation of coupled chemical-reactions. *J Phys Chem-Us* 81, 2340-2361.
- Grant, S.G., Jessee, J., Bloom, F.R., and Hanahan, D. (1990). Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proceedings of the National Academy of Sciences of the United States of America* 87, 4645-4649.
- Gregor, T., Tank, D.W., Wieschaus, E.F., and Bialek, W. (2007a). Probing the limits to positional information. *Cell* 130, 153-164.
- Gregor, T., Wieschaus, E.F., McGregor, A.P., Bialek, W., and Tank, D.W. (2007b). Stability and nuclear dynamics of the bicoid morphogen gradient. *Cell* 130, 141-152.
- Grenander, U. (1981). *Abstract Inference* (Wiley).
- Grinter, N.J. (1984). Replication control of IncP plasmids. *Plasmid* 11, 74-81.

- Guzman, L.M., Belin, D., Carson, M.J., and Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of bacteriology* 177, 4121-4130.
- Hammar, P., Leroy, P., Mahmutovic, A., Marklund, E.G., Berg, O.G., and Elf, J. (2012). The lac repressor displays facilitated diffusion in living cells. *Science* 336, 1595-1598.
- Hansen, A.S., and O'Shea, E.K. (2015). Limits on information transduction through amplitude and frequency regulation of transcription factor activity. *eLife* 4.
- Hao, N., and O'Shea, E.K. (2011). Signal-dependent dynamics of transcription factor translocation controls gene expression. *Nat Struct Mol Biol* 19, 31-39.
- Hasunuma, K., and Sekiguchi, M. (1977). Replication of plasmid pSC101 in *Escherichia coli* K12: requirement for dnaA function. *Mol Gen Genet* 154, 225-230.
- Helbl, V., and Hillen, W. (1998). Stepwise selection of TetR variants recognizing tet operator 4C with high affinity and specificity. *J Mol Biol* 276, 313-318.
- Hillen, W., Gatz, C., Altschmied, L., Schollmeier, K., and Meier, I. (1983). Control of Expression of the Tn10-Encoded Tetracycline Resistance Genes - Equilibrium and Kinetic Investigation of the Regulatory Reactions. *J Mol Biol* 169, 707-721.
- Hillen, W., Schollmeier, K., and Gatz, C. (1984). Control of Expression of the Tn10-Encoded Tetracycline Resistance Operon .2. Interaction of Rna-Polymerase and Tet Repressor with the Tet Operon Regulatory Region. *J Mol Biol* 172, 185-201.
- Hinrichs, W., Kisker, C., Duvel, M., Muller, A., Tovar, K., Hillen, W., and Saenger, W. (1994). Structure of the Tet repressor-tetracycline complex and regulation of antibiotic resistance. *Science* 264, 418-420.
- Howard, R.A. (1971). *Dynamic Probabilistic Systems: Markov Models: 1* (Dover Publications Inc).
- Jensen, J.L.W.V. (1906). On the convex functions and inequalities between mean values. *Acta Math-Djursholm* 30, 175-193.
- Johnson, J.E., Lackner, L.L., Hale, C.A., and de Boer, P.A. (2004). ZipA is required for targeting of DMinC/DicB, but not DMinC/MinD, complexes to septal ring assemblies in *Escherichia coli*. *Journal of bacteriology* 186, 2418-2429.
- Jorgensen, R.A., and Reznikoff, W.S. (1979). Organization of structural and regulatory genes that mediate tetracycline resistance in transposon Tn10. *Journal of bacteriology* 138, 705-714.
- Kaern, M., Elston, T.C., Blake, W.J., and Collins, J.J. (2005). Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* 6, 451-464.
- Kalir, S., McClure, J., Pabbaraju, K., Southward, C., Ronen, M., Leibler, S., Surette, M.G., and Alon, U. (2001). Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science* 292, 2080-2083.
- Kelly, J.R., Rubin, A.J., Davis, J.H., Ajo-Franklin, C.M., Cumbers, J., Czar, M.J., de Mora, K., Gliberman, A.L., Monie, D.D., and Endy, D. (2009). Measuring the activity of BioBrick promoters using an in vivo reference standard. *Journal of biological engineering* 3, 4.
- Khlebnikov, A., Datsenko, K.A., Skaug, T., Wanner, B.L., and Keasling, J.D. (2001). Homogeneous expression of the P(BAD) promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity AraE transporter. *Microbiology (Reading, England)* 147, 3241-3247.

- Kleinschmidt, C., Tovar, K., and Hillen, W. (1991). Computer simulations and experimental studies of gel mobility patterns for weak and strong non-cooperative protein binding to two targets on the same DNA: application to binding of tet repressor variants to multiple and single tet operator sites. *Nucleic Acids Res* 19, 1021-1028.
- Kleinschmidt, C., Tovar, K., Hillen, W., and Porschke, D. (1988). Dynamics of repressor-operator recognition: the Tn10-encoded tetracycline resistance control. *Biochemistry* 27, 1094-1104.
- Larrabee, K.L., Phillips, J.O., Williams, G.J., and Larrabee, A.R. (1980). The relative rates of protein synthesis and degradation in a growing culture of *Escherichia coli*. *J Biol Chem* 255, 4125-4130.
- Lau, B.T.C., Malkus, P., and Paulsson, J. (2013). New quantitative methods for measuring plasmid loss rates reveal unexpected stability. *Plasmid* 70, 353-361.
- Le, T.B., Schumacher, M.A., Lawson, D.M., Brennan, R.G., and Buttner, M.J. (2011). The crystal structure of the TetR family transcriptional repressor SimR bound to DNA and the role of a flexible N-terminal extension in minor groove binding. *Nucleic Acids Res* 39, 9433-9447.
- Lederer, T., Takahashi, M., and Hillen, W. (1995). Thermodynamic analysis of tetracycline-mediated induction of Tet repressor by a quantitative methylation protection assay. *Analytical biochemistry* 232, 190-196.
- Lestas, I., Vinnicombe, G., and Paulsson, J. (2010). Fundamental limits on the suppression of molecular fluctuations. *Nature* 467, 174-178.
- Lewin, B. (1994). *Genes V* (New York, Oxford University Press).
- Meier, I., Wray, L.V., and Hillen, W. (1988). Differential regulation of the Tn10-encoded tetracycline resistance genes *tetA* and *tetR* by the tandem tet operators O1 and O2. *Embo J* 7, 567-572.
- Miller, D.J., Zhang, Y.M., Subramanian, C., Rock, C.O., and White, S.W. (2010). Structural basis for the transcriptional regulation of membrane lipid homeostasis. *Nat Struct Mol Biol* 17, 971-975.
- Norman, T.M., Lord, N.D., Paulsson, J., and Losick, R. (2015). Stochastic Switching of Cell Fate in Microbes. *Annual review of microbiology* 69, 381-403.
- Orth, P., Schnappinger, D., Hillen, W., Saenger, W., and Hinrichs, W. (2000). Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat Struct Biol* 7, 215-219.
- Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D., and van Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. *Nat Genet* 31, 69-73.
- Palm, G.J., Lederer, T., Orth, P., Saenger, W., Takahashi, M., Hillen, W., and Hinrichs, W. (2008). Specific binding of divalent metal ions to tetracycline and to the Tet repressor/tetracycline complex. *J Biol Inorg Chem* 13, 1097-1110.
- Paninski, L. (2003). Estimation of entropy and mutual information. *Neural Comput* 15, 1191-1253.
- Paulsson, J. (2004). Summing up the noise in gene networks. *Nature* 427, 415-418.
- Pedraza, J.M., and van Oudenaarden, A. (2005). Noise propagation in gene networks. *Science* 307, 1965-1969.

- Perkins, T.J., and Swain, P.S. (2009). Strategies for cellular decision-making. *Mol Syst Biol* 5, 326.
- Purvis, J.E., Karhohs, K.W., Mock, C., Batchelor, E., Loewer, A., and Lahav, G. (2012). p53 dynamics control cell fate. *Science* 336, 1440-1444.
- Ramos, J.L., Martinez-Bueno, M., Molina-Henares, A.J., Teran, W., Watanabe, K., Zhang, X., Gallegos, M.T., Brennan, R., and Tobes, R. (2005). The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev* 69, 326-356.
- Raser, J.M., and O'Shea, E.K. (2005). Noise in gene expression: origins, consequences, and control. *Science* 309, 2010-2013.
- Rivoire, O., and Leibler, S. (2011). The value of information for populations in varying environments. *J Stat Phys* 142, 1124-1166.
- Ronen, M., Rosenberg, R., Shraiman, B.I., and Alon, U. (2002). Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proceedings of the National Academy of Sciences of the United States of America* 99, 10555-10560.
- Rosenfeld, N., Young, J.W., Alon, U., Swain, P.S., and Elowitz, M.B. (2005). Gene regulation at the single-cell level. *Science* 307, 1962-1965.
- Saenger, W., Orth, P., Kisker, C., Hillen, W., and Hinrichs, W. (2000). The Tetracycline Repressor-A Paradigm for a Biological Switch. *Angewandte Chemie (International ed)* 39, 2042-2052.
- Sanchez, A., and Golding, I. (2013). Genetic determinants and cellular constraints in noisy gene expression. *Science* 342, 1188-1193.
- Sawai, H., Yamanaka, M., Sugimoto, H., Shiro, Y., and Aono, S. (2012). Structural basis for the transcriptional regulation of heme homeostasis in *Lactococcus lactis*. *J Biol Chem* 287, 30755-30768.
- Schleif, R. (2000). Regulation of the L-arabinose operon of *Escherichia coli*. *Trends Genet* 16, 559-565.
- Schreiter, E.R., Wang, S.C., Zamble, D.B., and Drennan, C.L. (2006). NikR-operator complex structure and the mechanism of repressor activation by metal ions. *Proceedings of the National Academy of Sciences of the United States of America* 103, 13676-13681.
- Schultz, S.C., Shields, G.C., and Steitz, T.A. (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science* 253, 1001-1007.
- Selimkhanov, J., Taylor, B., Yao, J., Pilko, A., Albeck, J., Hoffmann, A., Tsimring, L., and Wollman, R. (2014). Accurate information transmission through dynamic biochemical signaling networks. *Science* 346, 1370-1373.
- Shannon, C.E. (1948). A mathematical theory of communication. *Bell Syst Tech J* 27, 379-423.
- Silva-Rocha, R., Martinez-Garcia, E., Calles, B., Chavarria, M., Arce-Rodriguez, A., de Las Heras, A., Paez-Espino, A.D., Durante-Rodriguez, G., Kim, J., Nikel, P.I., *et al.* (2013). The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res* 41, D666-675.

- Sliusarenko, O., Heinritz, J., Emonet, T., and Jacobs-Wagner, C. (2011). High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. *Mol Microbiol* 80, 612-627.
- So, L.H., Ghosh, A., Zong, C., Sepulveda, L.A., Segev, R., and Golding, I. (2011). General properties of transcriptional time series in *Escherichia coli*. *Nat Genet* 43, 554-560.
- Sotiropoulos, V., and Kaznessis, Y.N. (2007). Synthetic tetracycline-inducible regulatory networks: computer-aided design of dynamic phenotypes. *Bmc Syst Biol* 1, 7.
- Stanton, B.C., Nielsen, A.A., Tamsir, A., Clancy, K., Peterson, T., and Voigt, C.A. (2014). Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nature chemical biology* 10, 99-105.
- Sugino, Y., and Hirota, Y. (1962). Conjugal fertility associated with resistance factor R in *Escherichia coli*. *Journal of bacteriology* 84, 902-910.
- Swain, P.S., Elowitz, M.B., and Siggia, E.D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences of the United States of America* 99, 12795-12800.
- Takahashi, M., Altschmied, L., and Hillen, W. (1986). Kinetic and equilibrium characterization of the Tet repressor-tetracycline complex by fluorescence measurements. Evidence for divalent metal ion requirement and energy transfer. *J Mol Biol* 187, 341-348.
- Taniguchi, Y., Choi, P.J., Li, G.W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329, 533-538.
- Tkacik, G., Callan, C.G., Jr., and Bialek, W. (2008). Information flow and optimization in transcriptional regulation. *Proceedings of the National Academy of Sciences of the United States of America* 105, 12265-12270.
- Tkacik, G., Walczak, A.M., and Bialek, W. (2009). Optimizing information flow in small genetic networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 80, 031920.
- Van Kampen, N.G. (2007). *Stochastic Processes in Physics and Chemistry (NORTH-HOLLAND)*.
- Volkers, G., Petruschka, L., and Hinrichs, W. (2011). Recognition of drug degradation products by target proteins: isotetracycline binding to Tet repressor. *J Med Chem* 54, 5108-5115.
- Wissmann, A., Meier, I., Wray, L.V., Jr., Geissendorfer, M., and Hillen, W. (1986). Tn10 tet operator mutations affecting Tet repressor recognition. *Nucleic Acids Res* 14, 4253-4266.
- Wray, L.V., Jr., and Reznikoff, W.S. (1983). Identification of repressor binding sites controlling expression of tetracycline resistance encoded by Tn10. *Journal of bacteriology* 156, 1188-1191.
- Xu, Y., Heath, R.J., Li, Z., Rock, C.O., and White, S.W. (2001). The FadR-DNA complex. Transcriptional control of fatty acid metabolism in *Escherichia coli*. *J Biol Chem* 276, 17373-17379.
- Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M.G., and Alon, U. (2006). A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nature methods* 3, 623-628.