

Cada uno de los genes del genoma de un organismo son usados cuándo y dónde se los necesita.

Vimos el ejemplo del operón lac, que gobierna las enzimas responsables de la digestión de la lactosa en bacterias que enciende cuando hay lactosa pero no hay otras fuentes de glucosa. Algo similar ocurre con otros genes, tanto en bacterias como en células eucariotas.

Vamos a ver ahora cómo estudiar cuánto se expresa un gen determinado, en qué parte de la célula (o del organismo) se expresa ese gen y en qué momento durante la célula ciclo (o historia de vida) del organismo, para lo que usaremos algunas de las herramientas que fuimos viendo en el curso.

El enfoque de mecánica estadística se enfocará en la probabilidad de ocupación del promotor como la cantidad clave de interés, mientras que las ecuaciones que modelan la dinámica de las reacciones químicas serán usadas para estudiar las concentraciones de productos proteicos a lo largo del tiempo.

Estas mismas técnicas también se utilizarán para examinar la señalización con especial énfasis en las "decisiones" que toman las células sobre a dónde ir.

Las reacciones químicas que tienen lugar en las células están vinculadas entre sí en redes complejas

Vimos un ejemplo cuando analizamos brevemente la glucólisis, una parte pequeña de toda la red encarga del metabolismo de la glucólisis y la generación de ATP.

La bioquímica en las células vivas tiene reactivos y productos vinculados en un conjunto complejo en donde A engendra B, que engendra C, que a su vez engendra D, y así sucesivamente, con la complicación añadida de que Z podría simplemente engendrar A de nuevo.

Esta retroalimentación (la de Z engendrando A, por ejemplo) da lugar a una gran cantidad de comportamientos dinámicos diversos (e.g., las oscilaciones que vimos en el caso de la glucólisis).

Algo que distingue la dinámica de estas redes de la que se observa en solución donde solo está presente un subconjunto de las sustancias que participan de ellas es cómo, en la célula, mediante estos mecanismos de retroalimentación es posible ejercer un control que permite “sintonizar” las concentraciones en los valores requeridos para que la célula pueda funcionar correctamente.

Por otro lado, las reacciones que ocurren en las células también son estimuladas por señales externas dando lugar a las así llamadas **cascadas de señalización**.

Las reacciones químicas que tienen lugar en las células están vinculadas entre sí en redes complejas

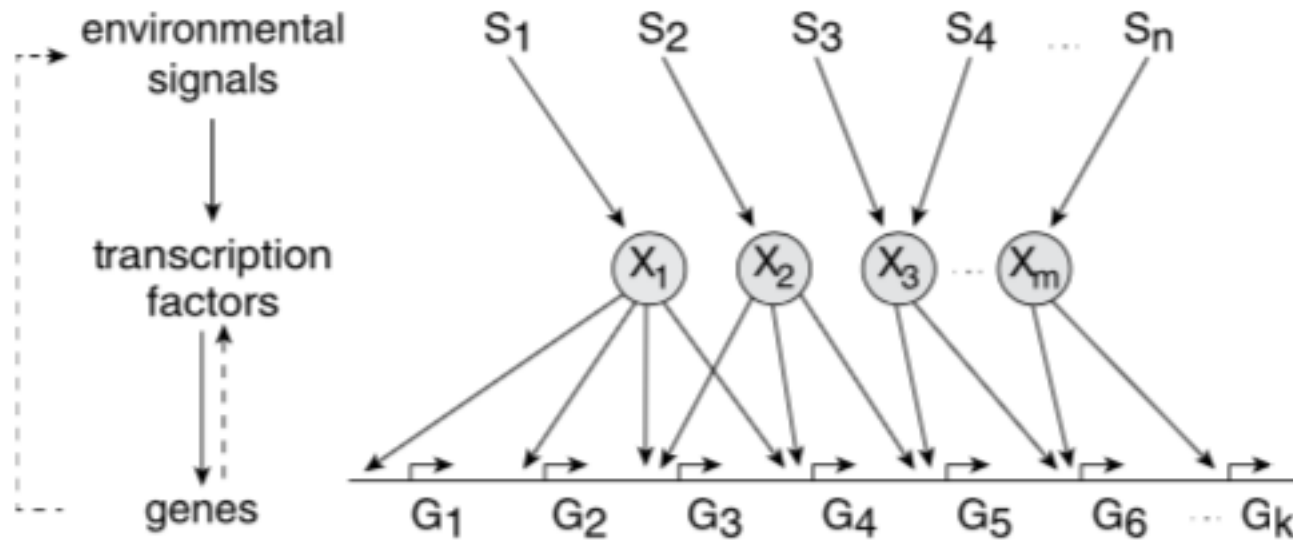
Vimos un ejemplo cuando analizamos brevemente la glucólisis, una parte pequeña de toda la red encarga del metabolismo de la glucólisis y la generación de ATP.

La bioquímica en las células vivas tiene reactivos y productos vinculados en un conjunto complejo en donde A engendra B, que engendra C, que a su vez engendra D, y así sucesivamente, con la complicación añadida de que Z podría simplemente engendrar A de nuevo.

Esta retroalimentación (la de Z engendrando A, por ejemplo) da lugar a una gran cantidad de comportamientos dinámicos diversos (e.g., las oscilaciones que vimos en el caso de la glucólisis).

Algo que distingue la dinámica de estas redes de la que se observa en solución donde solo está presente un subconjunto de las sustancias que participan de ellas es cómo, en la célula, mediante estos mecanismos de retroalimentación es posible ejercer un control que permite “sintonizar” las concentraciones en los valores requeridos para que la célula pueda funcionar correctamente.

Por otro lado, las reacciones que ocurren en las células también son estimuladas por señales externas dando lugar a las así llamadas **cascadas de señalización**.



Esquema del libro “An Introduction of Systems Biology” de Uri Alon

Los cambios en el entorno son “leídos” por receptores ubicados en la membrana plasmática. Estos inducen cambios en el interior celular que, en algunos casos, dan lugar a cambios en la expresión génica.

En algunos casos no solo se pueden combinar varios “estímulos externos” para generar la respuesta, sino que ésta depende también del estado particular en el que se encuentra la célula en ese momento (de su estado interno). O sea, las reacciones están acopladas de un modo complejo.

Un ejemplo: el apareamiento de las células de levadura (organismos modelo en el que trabaja Alejandro Colman-Lerner)

Cuando los nutrientes son suficientes, las células de levadura se reproducen simplemente por división celular.

Cuando los nutrientes empiezan a escasear aparecen dos sub-tipos (a y α , genotipos MAT a y MAT α , respectivamente) que pueden aparearse entre sí para dar lugar a una célula hija.

Las células MAT α secretan la feromona, factor α (un péptido de 13 residuos de secuencia WHWLQLKPGQPMY) y responden a la feromona, factor a . Las células MAT a secretan factor a (un péptido de 12 residuos de secuencia YIIKGVFWDPAC) y responden al factor α .

Cuando una célula de levadura es estimulada por la feromona secretada por una célula cercana del tipo de apareamiento opuesto, sufre una serie de cambios fisiológicos en preparación para el apareamiento.

Estos incluyen cambios significativos en la expresión de aproximadamente 200 genes (~3% del genoma), se “arrestan” (detienen) en la fase G1 del ciclo celular y comienzan a crecer (a desarrollar un “shmoo”) en la dirección de la pareja de apareamiento.

Finalmente, las dos células de la pareja entran en contacto, fusionan sus membranas plasmáticas y, poco después sus núcleos. Todo este proceso dura unas 4 horas.

Los procesos que ocurren en la célula MAT a cuando las células tipo MAT α secretan el factor α se inducen experimentalmente agregando factor α (purificado) en el baño.

Expresando versiones fluorescentes de algunas proteínas involucradas en los cambios es posible estudiar las cascadas de señalización a las que da lugar esta feromona.

La vía de transducción de señales que detecta la presencia de feromona extracelular y orquesta las respuestas celulares correspondientes es conocida como **vía de respuesta de apareamiento a feromonas**.

El primer paso consiste en la unión de la feromona de apareamiento a un receptor acoplado a proteína G (GPCR). La ocupación del receptor estimula a la subunidad $G\alpha$ de la proteína G a intercambiar GDP por GTP. $G\alpha$ unido a GTP puede, entre otras cosas, liberar el heterodímero $G\beta\gamma$.

El flujo de información procede entonces de $G\beta\gamma$ a través de una cascada de proteína quinasa de cuatro niveles a factores de transcripción nuclear y otros objetivos. Los componentes principales de la vía y sus funciones se resumen en la Tabla.

Table 1
Some key components of the yeast mating pheromone response pathway

Protein	Function
Ste2/3	7-transmembrane-segment, G-protein coupled pheromone receptors
Gpa1	G-protein α subunit
Ste4, Ste18	G-protein $\beta\gamma$ subunits
Ste5	Adapter and scaffold, binds $G\beta$, MAPK cascade kinases, and others
Bem1	Involved in polarity establishment, binds Ste5, Cdc42, Cdc24 and Ste20
Cdc24	Guanine nucleotide exchange factor (GEF) for Cdc42
Cdc42	Small rho-like G-protein, binds to Ste20, Bem1, and others
Ste20	PAK (p21-activated protein kinase), activated by Cdc42
Ste11	MEKK (MEK kinase), activated by Ste20
Ste50	Binds to N-terminus of Ste11 and aids and/or helps maintain in its activation
Ste7	MEK (MAPK/ERK kinase), activated by Ste11
Kss1, Fus3	MAP kinases, activated by Ste7
Dig1, Dig2	MAPK substrates, repressors of Ste12 transcriptional activity
Ste12	MAPK substrate, DNA-binding transcriptional transactivator
Far1	MAPK substrate, inhibits cell-cycle progression, also adapter/scaffold that binds $G\beta$, Cdc24 and others

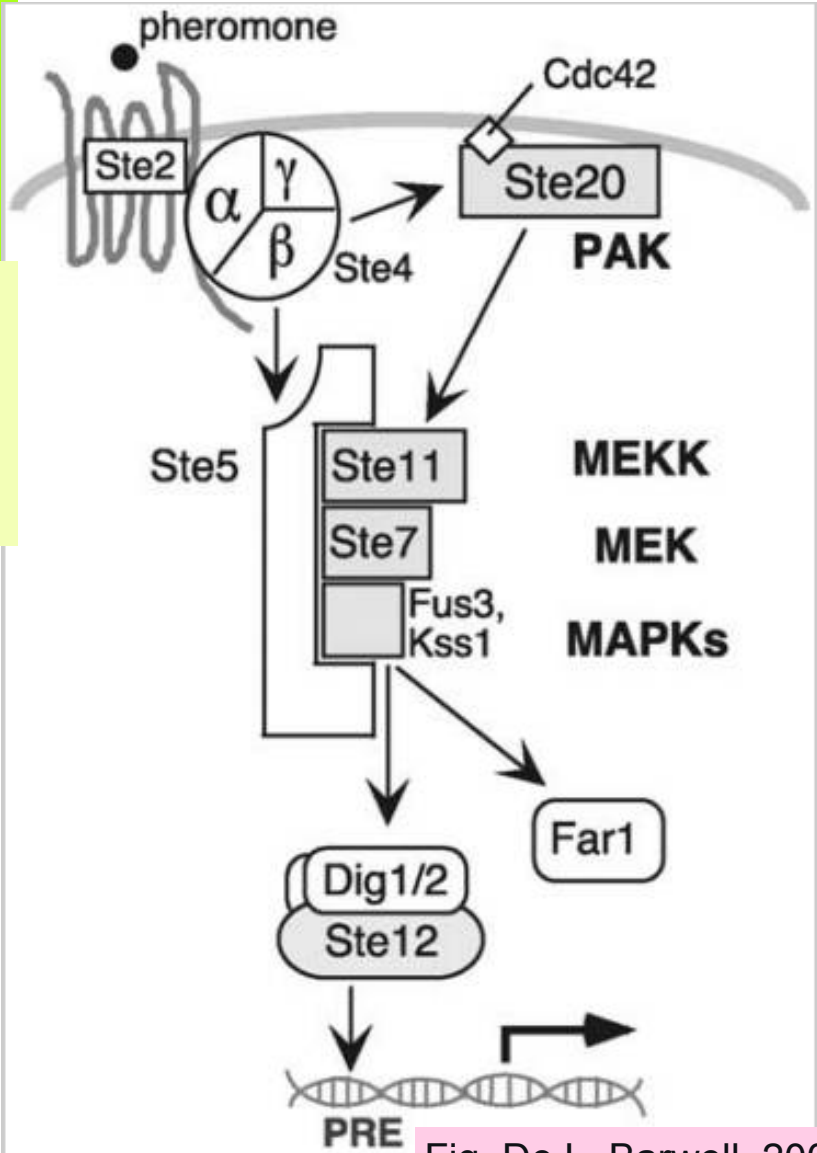


Fig. De L. Barwell, 2005

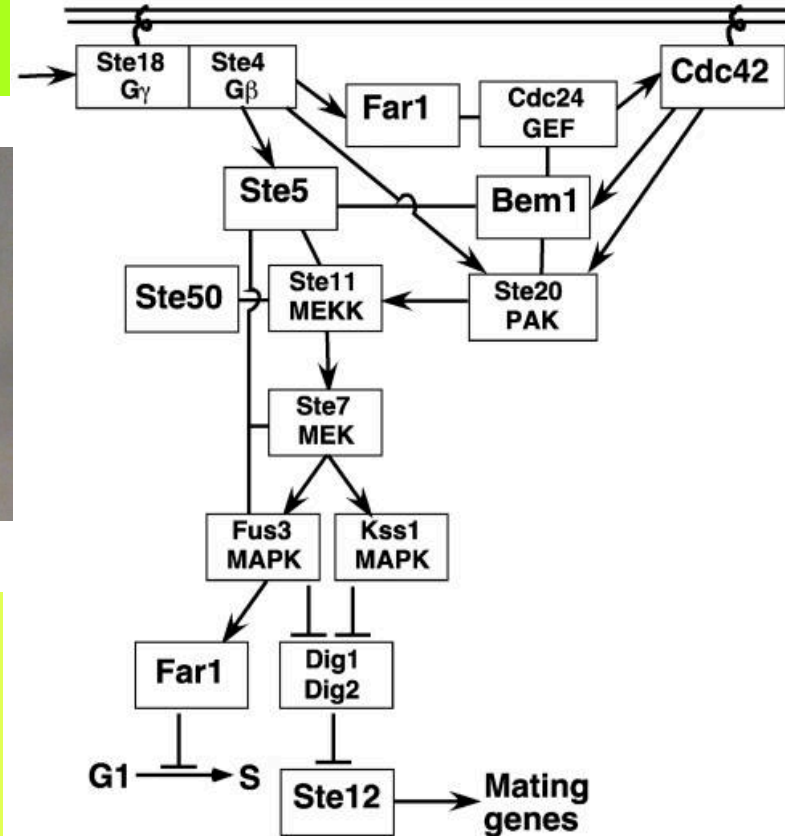
Vemos que es toda una secuencia compleja de reacciones que interactúan entre sí, incluso el trabajo (de hace varios años) del que saqué la figura, tiene otra con más detalles:

En el caso de la levadura, como las células no se desplazan, estos “mating genes” lo que hacen es que la célula cambie su forma (y desarrolle el “shmoo”).

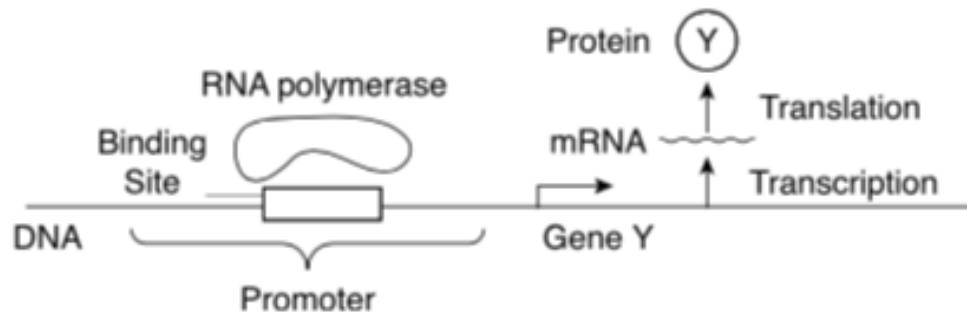


Éste es un crecimiento asimétrico, en una dirección particular, por lo que se dice que está polarizado, y como la dirección elegida es hacia la concentración más alta de feromona, se dice además que es quimiotrópico. La rama $G\beta$ -Far1-Cdc24-Cdc42 de la vía es crucial en este proceso. Las células que se desplazan utilizan estrategias reguladoras similares por ejemplo, el reclutamiento dependiente de $G\beta\gamma$ de un cierto factor también ocurre en la quimiotaxis en mamíferos.

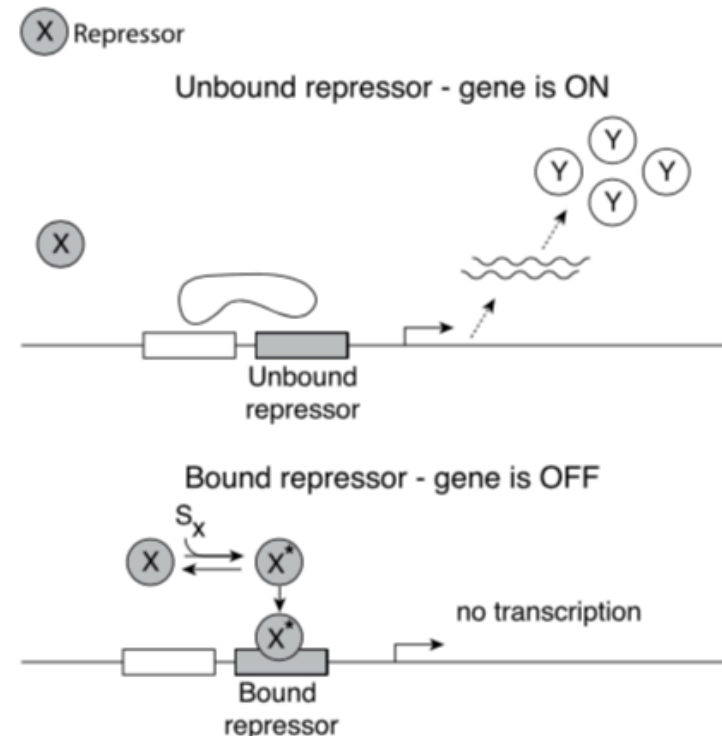
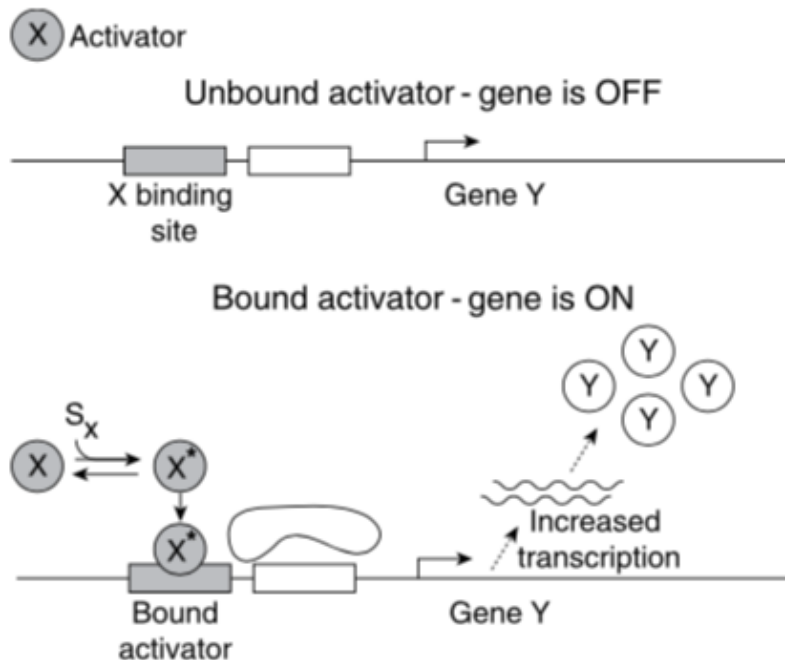
Fig. De L. Barwell, 2005



En este caso (y en muchos otros), parte de la red involucra inducir cambios en la expresión génica. Es decir, variar la concentración (dentro del núcleo) de factores de transcripción específicos. Recordemos los pasos que van de la lectura del ADN a las proteínas

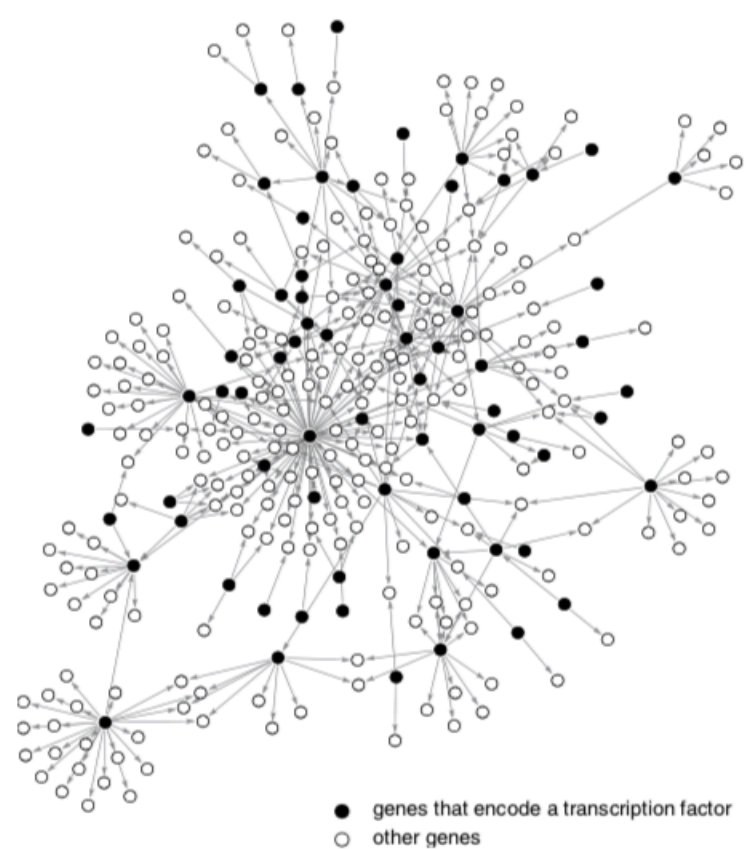


Y algunas formas en las que se puede regular la expresión génica:



Figs. De "An Introduction of Systems Biology" de Uri Alon

Los factores de transcripción, por otro lado, son típicamente proteínas que están codificadas en sí mismas por genes, que están regulados por otros factores de transcripción, que a su vez pueden estar regulados por otros factores de transcripción, y así sucesivamente. Este conjunto de interacciones forma una red de transcripción. En la figura (del libro de U. Alon), los nodos son genes y las flechas representan la regulación transcripcional de un gen por el producto proteico de otro gen. Una flecha $X \rightarrow Y$ significa que el producto del gen X es una proteína de factor de transcripción que puede unirse al promotor del gen Y para controlar la velocidad a la que se transcribe el gen Y.



En los diagramas de redes también se representa la naturaleza de la acción (activación o represión) que ejerce un componente sobre otro. La activación se indica con una flecha regular $X \rightarrow Y$, mientras que la represión se indica con una flecha de punta roma $X \dashrightarrow Y$. Las redes de transcripción a menudo muestran números comparables de flechas positivas y negativas, con más interacciones positivas (activación) que interacciones negativas (e.g., aproximadamente un 60% de activación en *E. coli* y un 80% en levadura).

Cada factor de transcripción actúa principalmente de un dado modo (activador o represor) sobre los genes que regula. Los modos de regulación de entrada a menudo se mezclan: un gen típico es activado por algunos factores de transcripción y reprimido por otros. Los signos de las flechas de salida están muy correlacionados, pero los de las de entrada no.

La química celular está bajo el estricto control de la maquinaria genética de la célula. Si bien algunas vías químicas pueden tener una dinámica que las lleva a algún equilibrio, las señales externas o internas pueden llevar a la expresión de algún gen que da como resultado un nuevo reactivo en la vía química original que lo altera el equilibrio.

Esto sucede tanto para responder a cambios en el ambiente como durante el desarrollo de los organismos. Por ejemplo, la figura muestra una red genética muy bien caracterizada que participa en el desarrollo embrionario de los erizos de mar.

Esto da una idea de la complejidad de estas redes.

Las redes genéticas como la que se muestra en la figura no hacen referencia ni al paso del tiempo ni a las distribuciones cuantitativas de las moléculas que las median, ni a la distribución espacial dentro de la célula de las sustancias involucradas. Son más bien una abstracción que muestra cómo los genes (y sus productos) están vinculados entre sí.

Debajo de la superficie de estos diagramas de cableado se encuentran las concentraciones reales con sus dependencias espaciales y temporales de los jugadores moleculares de involucrados en las vías.

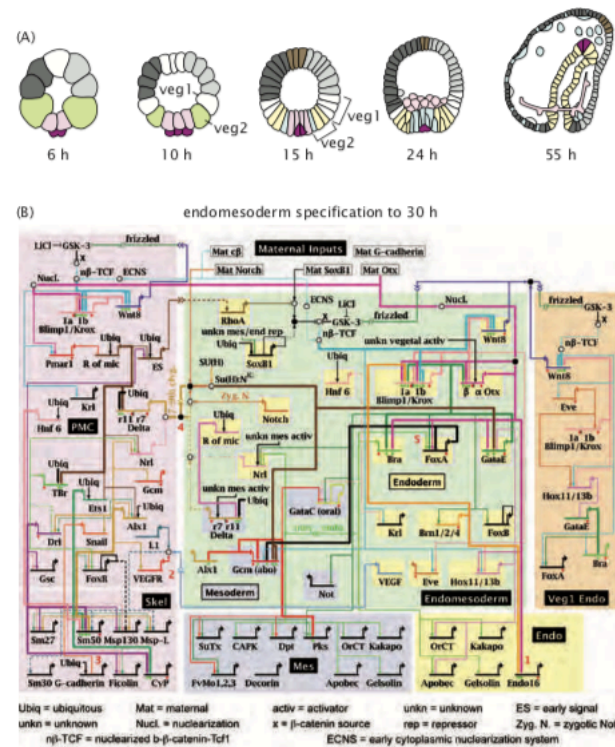


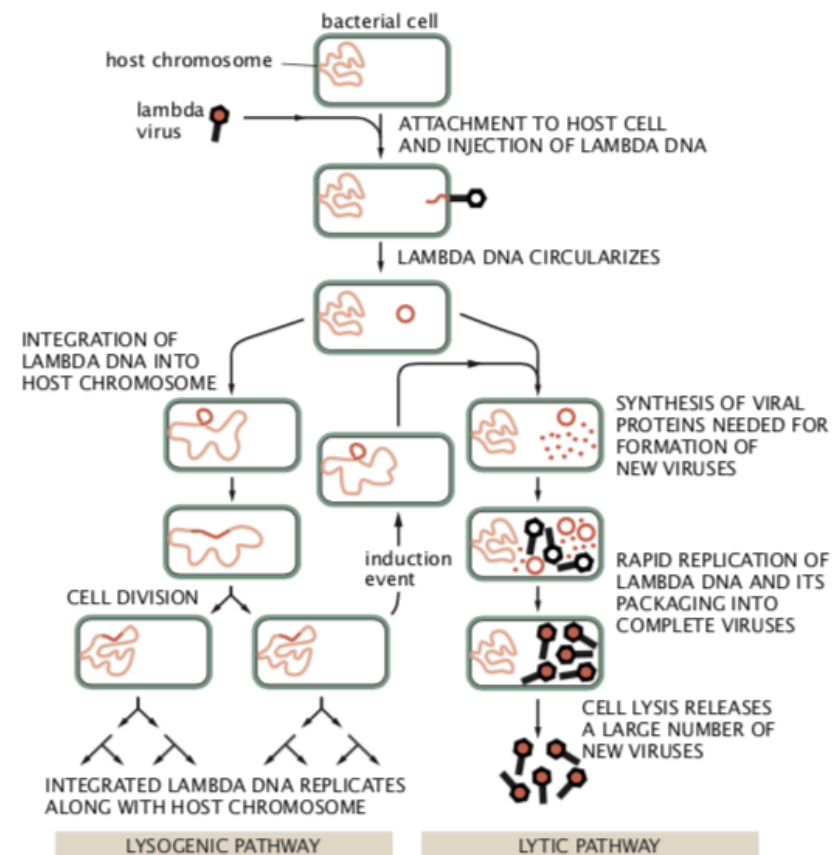
Figure 19.1: Genetic network associated with control of the developmental pathway of the sea urchin embryo. (A) Schematic of stages in the embryonic development of the sea urchin. (B) Genetic network associated with sea urchin development. (Adapted from S. Ben-Tabou de-Leon and E. H. Davidson, *Annu. Rev. Biophys. Biomol. Struct.* 36:191, 2007.)

El desarrollo embrionario involucra la regulación de genes

Todas las células de un organismo multicelular llevan la misma información genética. Sin embargo, en general, las células solo expresan una cierta fracción de todos los genes disponibles. Esta diferenciación es la base del desarrollo de los embriones y la base de las diferentes estructuras que se encuentran en los organismos multicelulares. El punto clave es que no todos los genes se expresan todo el tiempo.

Uno de los ejemplos más famosos de una "decisión de desarrollo" es el interruptor lambda. Después de infectar una bacteria *E. coli*, el fago λ sigue una de dos vías posibles de desarrollo. Una de ellas, la vía lítica, da como resultado el ensamblaje de nuevos fagos y la lisis de la célula huésped. La segunda vía, la vía lisogénica, implica la incorporación del genoma del λ al de la célula huésped. La lisogenia se puede revertir al dañar la célula con luz ultravioleta, lo que desencadena la replicación lítica.

Figure 4.10: Lifestyles of bacterial viruses. Lysogenic and lytic pathways for bacteriophage lambda. In the lysogenic pathway, the viral genome is incorporated into the bacterial genome to form a "prophage." In the lytic pathway, new viruses are assembled to completion and are able to repeat their deeds elsewhere after being released from the lysed cell. (Adapted from B. Alberts et al., *Molecular Biology of the Cell*, 5th ed. Garland Science, 2008.)



La vida de los bacteriófagos (virus que infectan bacterias)

Los distintos procesos que ocurren cuando los bacteriófagos infectan bacterias son: la infección que consiste en el ingreso del ADN viral en la célula huésped; la transcripción y traducción de la información genética de este ADN que ocurren cuando la maquinaria celular es “secuestrada” para producir componentes virales (tanto ácido nucleico como proteínas); el ensamblaje que es la unión de estos bloques de construcción para formar la cápside viral y el empaquetamiento que es la parte del ciclo cuando el genoma viral está encerrado dentro de la cápside. Una de las consecuencias de esta infección es la lisis, es decir, la disolución de la célula huésped. Esto da lugar a una nueva generación de fagos que repiten su ciclo de vida en otros lugares. Todo esto ocurre en aproximadamente 30 minutos.

Todo esto ocurre en aproximadamente 30 minutos.

El represor del fago λ es la proteína de unión al ADN que, en un lisógeno, desactiva la transcripción de los genes del fago lítico. El represor también controla la expresión de su propio gen (*cI*), tanto positiva como negativamente. La combinación de estos efectos mantiene la concentración de represor en el nivel correcto como para responder rápidamente a una señal inductora como la irradiación UV. Ésta induce el crecimiento del fago lítico al destruir el represor. Cuando los genes líticos se activan, el gen represor se desactiva. El switch es muy eficaz: en ausencia de agente inductor, un lisógeno es estable durante incontables generaciones pero al recibir la señal, en prácticamente todas las células de la población se produce un crecimiento lítico del fago previamente inactivo

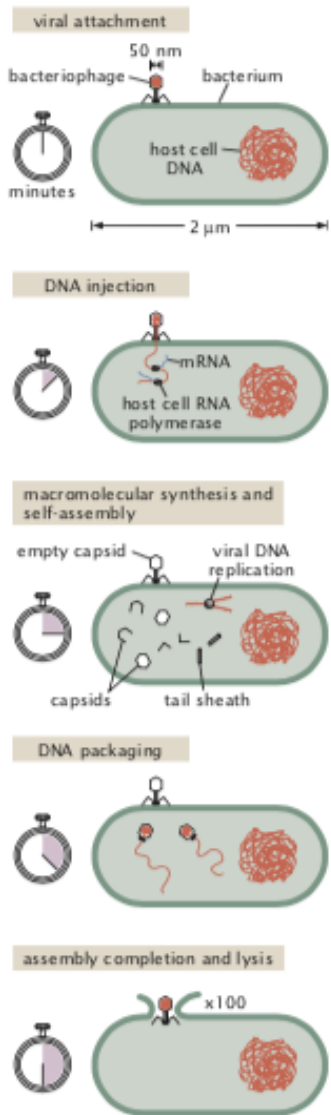
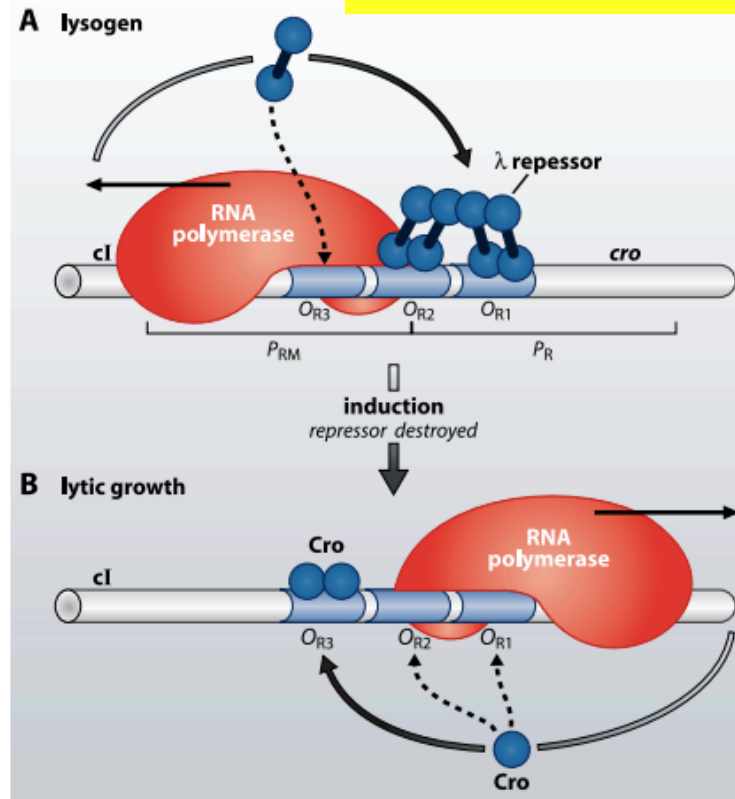


Figure 3.26: Timing the life cycle of a bacteriophage. The cartoon shows stages in the life cycle of a bacteriophage and roughly how long after infection these processes occur. Note that the head and the tail follow distinct assembly pathways and join only after they have separately assembled.

Un poco más de detalle.

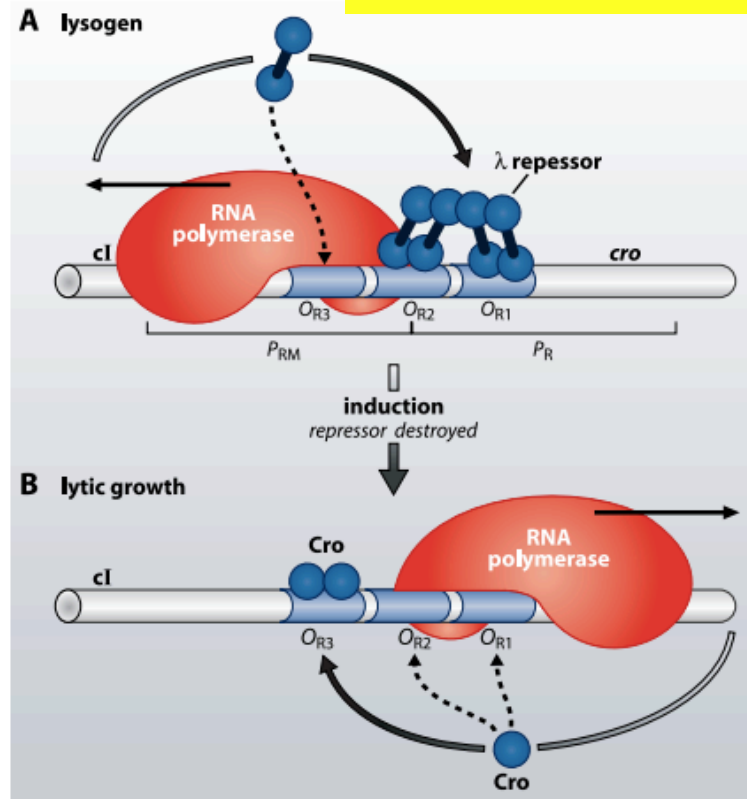


En el ADN hay dos genes adyacentes, uno que se transcribe a izquierda y otro a la derecha. El que se transcribe a izquierda, *cI*, está involucrado en la producción del represor λ y está regulado por el propio represor, por lo que es el propio represor el que mantiene su concentración dentro de cierto rango. El gen que se transcribe a derecha (*cro*) es el de la proteína Cro, que se encuentra silente en un lisógeno, es uno de los primeros en transcribirse tras la inducción, y su actividad es necesaria para un crecimiento lítico eficaz.

(A) Dos dímeros represores se unen cooperativamente a los sitios O_{R1} y O_{R2} del operador a derecha, O_R , del ADN de un lisógeno (flechas sólidas). A través de un contacto proteína-proteína, el represor en O_{R2} recluta a la ARN polimerasa hacia el promotor (P_{RM}) del gen adyacente, *cI*, que es el gen del represor del fago λ . Los represores ligados también excluyen a la polimerasa de otros promotores que están a los costados (P_R), manteniendo así a los genes líticos desactivados.

Con una afinidad más baja, el represor se une a O_{R3} (flecha discontinua) y desactiva la transcripción de *cI* y así regula la concentración del represor. Un segundo promotor de los genes líticos (P_L) que está a unos 2400 pares de bases de distancia y no se muestra en la figura contribuye a que ocurran estas reacciones (especialmente la unión del represor a O_{R3} gracias a que se hace un "loop" en el ADN).

Un poco más de detalle.



(B) Si se induce la destrucción del represor, la transcripción de *cI* (el gen del represor) cae (ya que el represor está involucrado en su expresión) y se produce la transcripción hacia la derecha de genes líticos, en particular, de *cro*.

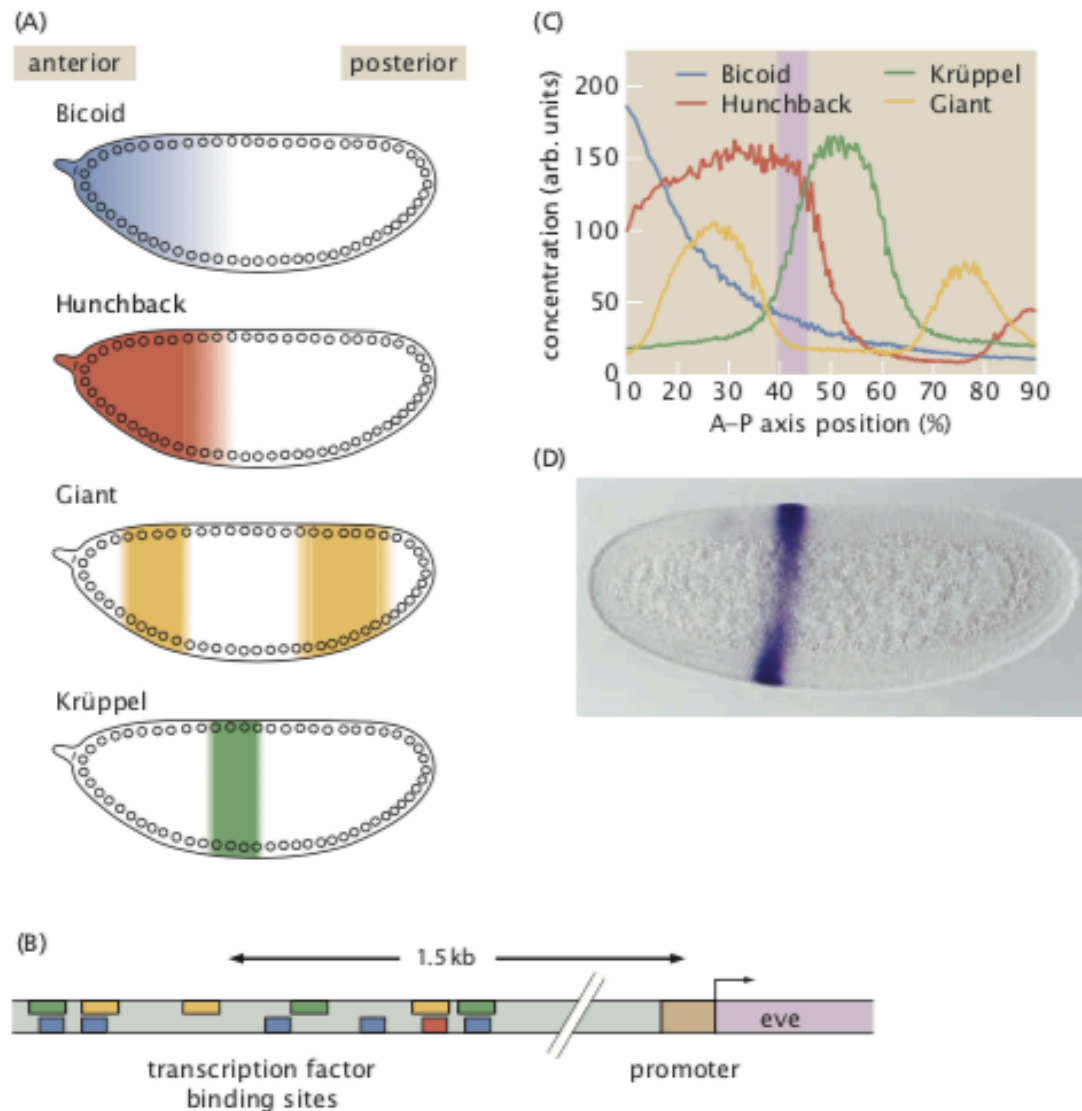
Una función esencial de la proteína, Cro, es disminuir la expresión de genes líticos (incluido él mismo) que se expresan en un nivel alto inmediatamente después de la inducción. Lo hace uniéndose a O_{R1} y O_{R2} , los mismos sitios reconocidos por el represor.

Por otro lado, entre los genes que desactiva Cro se encuentra uno que codifica a la proteína, CII, que es la proteína que inicia la transcripción de *cI*, que luego se vuelve autosostenida. Por tanto, Cro desalienta indirectamente el establecimiento de la síntesis de represores y por lo tanto estimula al fago a entrar en el ciclo lítico después de la irradiación UV de un lisógeno lambda.

Volviendo a la figura, se supone que Cro se une con más afinidad a O_{R3} (flecha sólida), reprimiendo así directamente la síntesis represora. Eventualmente, Cro se une a los sitios O_{R2} y O_{R1} y apaga la transcripción de genes tempranos (líneas discontinuas). Entre los genes que se apagan de este modo están el propio Cro y *cII*, un activador transcripcional que, para establecer la lisogenia, inicia la transcripción transitoria de *cI* a partir de un promotor que está a la derecha de Cro (que no se muestra en la fig.).

Otro ejemplo es el desarrollo embrionario en las moscas de la fruta (del que ya hablamos muchas veces), en particular, el del establecimiento del eje antero-posterior. En la figura se ven los gradientes en 4 proteínas reguladoras clave que determinan el patrón de expresión génica a lo largo del embrión, del cual la franja Eve 2 es el ejemplo mejor entendido.

Figure 19.2: Regulatory proteins in the *Drosophila* embryo. The anterior-posterior (A-P) patterning of the fruit fly is dictated by genes that are controlled by spatially varying concentrations of transcription factors. (A) Schematic of the main transcription factors involved in the regulation of stripe 2 of expression of the *even-skipped* gene (*eve*). (B) Regulatory region of the stripe 2 of the *even-skipped* gene where the binding sites for each transcription factor have been identified. The binding site color on the DNA corresponds to the transcription factor color in (A). (C) Spatial profile of the morphogen gradients measured using immunofluorescence. The purple shaded region corresponds to the striped region shown in (D). (D) Resulting pattern of expression of the regulatory region shown in (B). (B, Adapted from S. Small et al., *EMBO J.* 11:4047, 1992.; C, adapted from E. Myasnikova et al., *Bioinformatics* 17:3, 2001; D, adapted from S. Small et al., *Dev. Biol.* 175:314, 1996.)



En clases anteriores mencionamos que hay muchas etapas en la ruta entre el ADN y la proteína funcional que pueden servir como puntos reguladores de la cantidad de proteínas de distintos tipos que existen en las células y que uno de los mecanismos reguladores más comunes es el control transcripcional, donde la decisión clave que se toma es si producir o no ARNm.

La expresión génica es un tema que se ha vuelto cada vez más cuantitativo. En particular, ahora es común medir cuánto se expresa un gen determinado, cuándo se expresa y dónde se expresa. Para realizar tales mediciones, existen varias herramientas útiles.

Describimos a continuación algunos ejemplos que sirven a la cuantificación.

Una forma importante de caracterizar la actividad de un gen es en virtud de sus productos proteicos. En particular, si el producto génico tiene actividad enzimática, esa actividad se puede analizar como un indicador del grado en que se ha expresado el gen. El ejemplo de la figura corresponde a la β -galactosidasa, el producto enzimático del operón lac cuya acción consiste en recortar las moléculas de lactosa.

Existe un sustrato (ONPG) que es cortado por esta enzima que se vuelve amarillo tras el corte. Esto permite cuantificar la cantidad de enzima midiendo la velocidad a la que una solución se vuelve amarilla ópticamente.

Estas mediciones se realizan normalmente en poblaciones de células y requieren lisar las células: solo se pueden realizar ensayos de punto final con esta técnica. La sensibilidad del método es excelente: se puede medir la actividad de menos de una molécula de β -galactosidasa por célula. Para hacer estos ensayos se requiere una clonación rutinaria en la que las secuencias que codifican la enzima se insertan en el genoma bajo el control de los factores de transcripción de interés.

También se puede insertar un informador en el gen de interés de modo que la molécula "informadora" sea fluorescente (e.g., GFP) en lugar de una enzima como en el caso anterior. La GFP permite monitorear el nivel de expresión génica en función del tiempo en células vivas individuales. Sus desventajas son que las moléculas pueden sufrir fotoblanqueo y que las células tienen una fluorescencia intrínseca, lo que da como resultado un fondo de autofluorescencia celular que potencialmente puede contaminar la lectura del reportero de GFP.

Esta figura ilustra el uso del análisis de la actividad enzimática o de un reportero fluorescente directamente para analizar expresión génica.

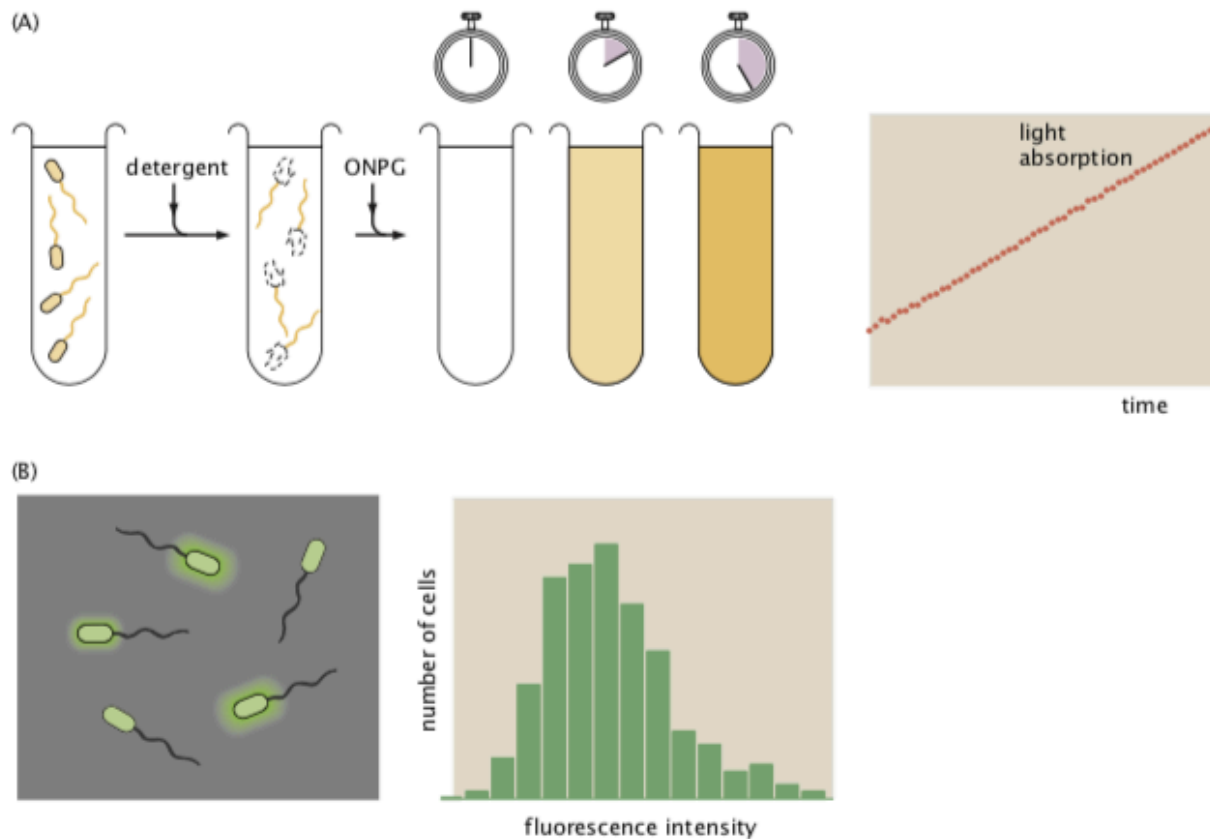


Figure 19.3: Measurement of gene expression. (A) Measurement of gene expression as a result of enzymatic activity. The promoter of interest drives the expression of an enzyme that can cleave a molecule that in the cleaved state is colored. The resulting rate of increase in light absorption is related to the amount of enzyme present in the cells. (B) The promoter of interest drives the expression of a fluorescent protein such as GFP. The amount of fluorescence per cell reports the extent of expression of the gene of interest.

Un segundo esquema para caracterizar el grado en que se expresa un gen es midiendo cuánto mRNA está presente en la célula. Una herramienta posible lo constituyen los microarrays de ADN que se construyen marcando una superficie con una matriz de diferentes moléculas de ADN, cada parche de los cuales tiene moléculas pequeñas de ADN monocatenario (ssDNA) con la misma secuencia. Estas secuencias se eligen para complementar una batería completa de secuencias correspondientes a los genes de interés en el experimento. Luego, las células se rompen y se permite que su ARN (o copias de ADN hechas a partir del ARN) fluya a través de la matriz y se hibridice con las moléculas en la superficie. Marcando fluorescentemente las moléculas de interés en la célula midiendo la fluorescencia en cada punto del array es posible leer la cantidad de ARN presente.

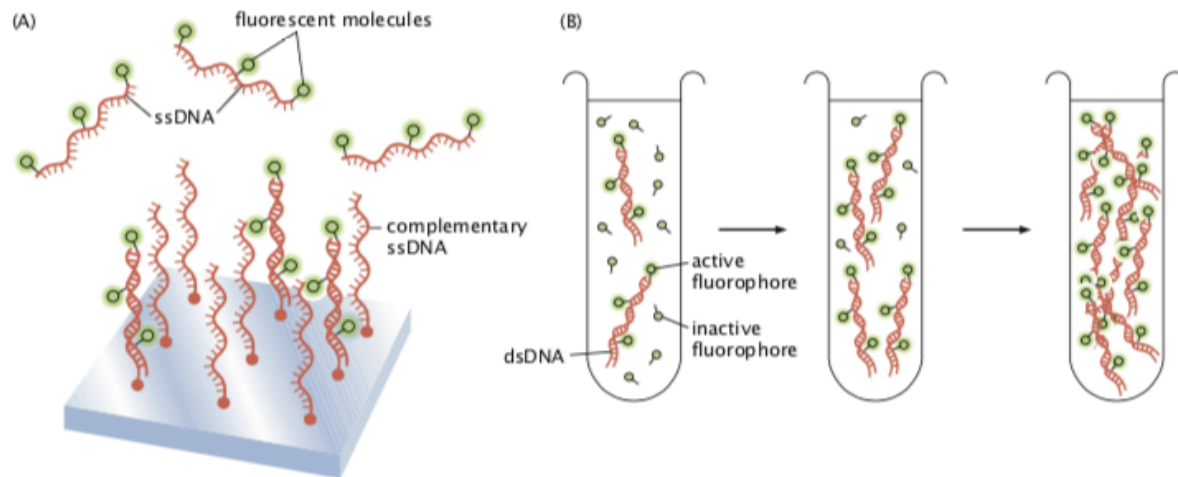


Figure 19.4: Measurement of mRNA concentration. (A) A DNA microarray uses a collection of different molecules on the surface of a slide, each of which has a sequence complementary to the mRNA (or reverse-transcribed ssDNA) associated with the gene of interest. By measuring how much hybridization there is between the sample and the molecules on the surface, one can count the mRNAs. (B) Quantitative PCR uses a template molecule that is produced from the mRNA using reverse transcription. The amount of template determines how many cycles of PCR it will take to reach a critical threshold of amplified DNA using fluorescence as a readout.

Otro esquema para caracterizar la cantidad de ARN es utilizar PCR cuantitativa. Para ello se lisan las células y las moléculas de ARNm se convierten en ADN mediante una reacción de transcripción inversa. Luego, estas moléculas se utilizan como plantillas en una PCR para estimar si están o no presentes y en qué cantidad.

Ahora es posible observar la transcripción en tiempo real usando sistema de imágenes MS2/“MS2 coat protein”-GFP (MCP-GFP)

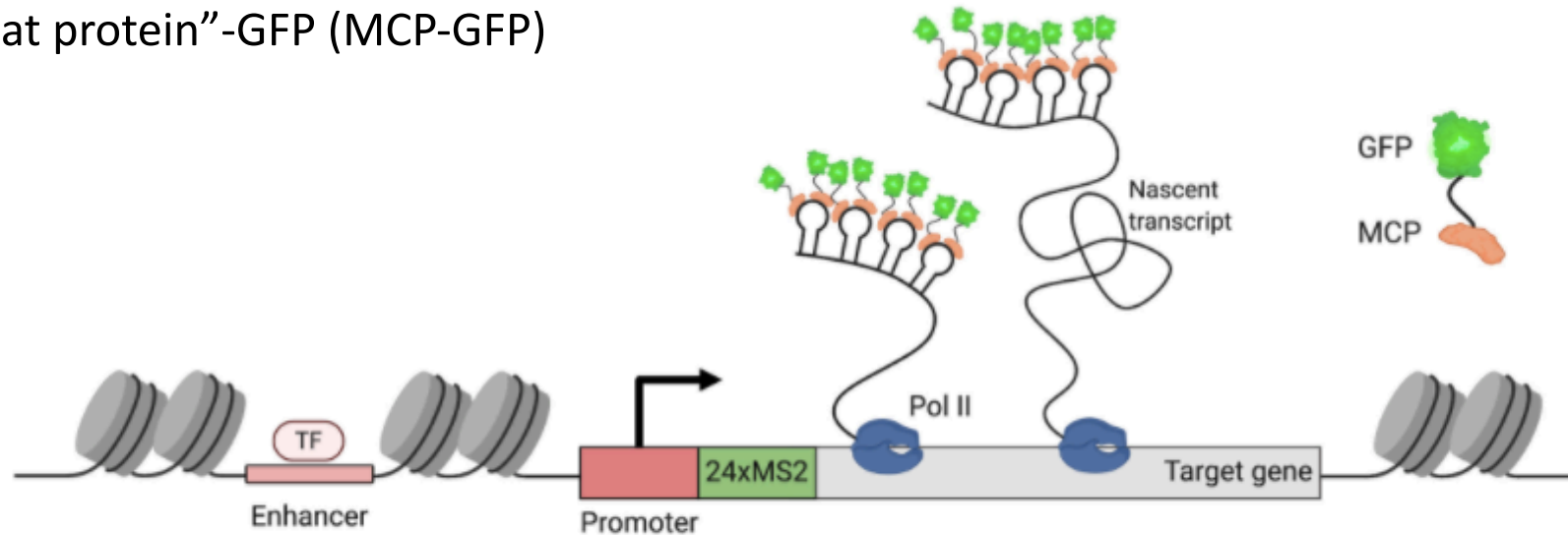


Figure 1. Visualization of transcription with the MS2/MCP imaging system

The MS2 live imaging system uses RNA stem-loop repeats, derived from the MS2 bacteriophage that are bound by the cognate MS2-coat protein (MCP) fused to a fluorescent protein. Transcription factor = TF.

Este sistema consiste en insertar en un lugar adecuado del ADN la secuencia derivada del bacteriófago MS2 que está asociada a la producción de un “stem-loop” en el ARN. Estas secuencias se insertan en regiones no codificantes de genes endógenos o de transgenes del organismo/célula a estudiar. Una vez transcrito por la ARN polimerasa II (Pol II), el “stem-loop” es reconocido por la proteína, MCP, que se une a la estructura del loop como un dímero. La idea es usar MCP con una cola fluorescente (GFP). De este modo, a medida que se va produciendo el ARNm va aumentando la fluorescencia. Cuantificando dicha fluorescencia es posible cuantificar la cantidad de ARNm que se va produciendo. El sistema MCP fue usado para observar la dinámica de la transcripción a nivel de célula única con alta resolución espacial y temporal tanto en células en cultivo como *in vivo*.

En la figura se ilustran dos métodos complementarios para estudiar los niveles de expresión en *Drosophila* (Gregor et al, 2014): uno que permite obtener imágenes en embriones fijados y el otro, in vivo. En el primer método, se marcan proteínas o moléculas de ARNm en embriones fijados químicamente para proporcionar una instantánea muy precisa del sistema intacto. Aunque requiere detener el proceso de desarrollo, este enfoque tiene la ventaja de que se puede aplicar fácilmente siempre que se disponga de reactivos de marcado adecuados. El segundo método permite la observación directa de la dinámica marcando con indicadores fluorescentes codificados genéticamente. En estos casos, los embriones se pueden observar en tiempo real a medida que se produce el ARNm. Sin embargo, se debe tener mucho cuidado para determinar si dicho etiquetado altera o no las características del proceso de desarrollo que podría verse perturbado por la adición de indicadores sintéticos.

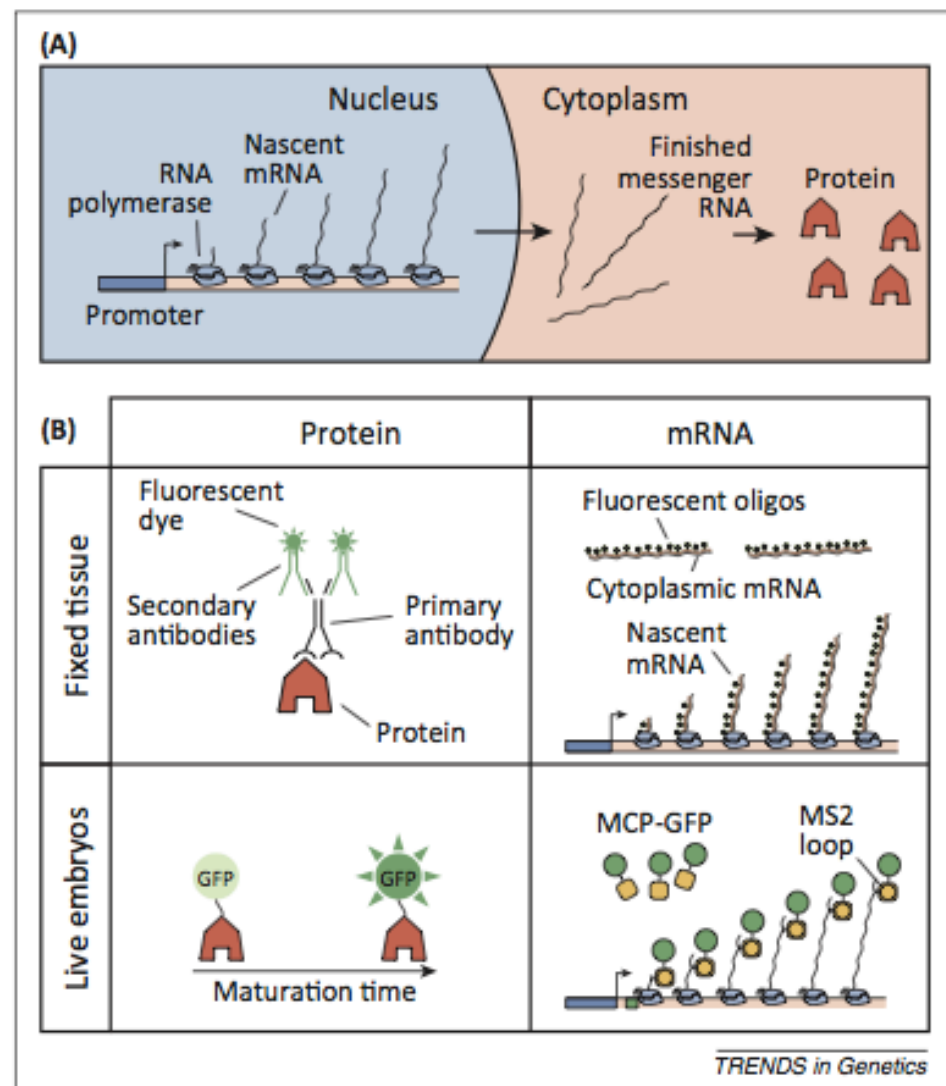


Figure 2. Visualizing the central dogma in the embryo. **(A)** Developmental decisions can be assessed at multiple stages of the central dogma: from nascent mRNA transcript formation (and thus the number of active RNA polymerases loaded on the promoter) to cytoplasmic mRNA and protein distributions. **(B)** Methods to measure mRNA and protein distributions during development in fixed and living embryos (see text).

En los modelos "termodinámicos" de expresión génica, la atención se centra en la probabilidad de que el promotor esté ocupado por la ARN polimerasa. Ahora bien, esta probabilidad de ocupación depende también de la interacción de entre los promotores y los factores de transcripción (efecto que no tuvimos en cuenta cuando hicimos algunos cálculos de funciones de partición hace unas clases).

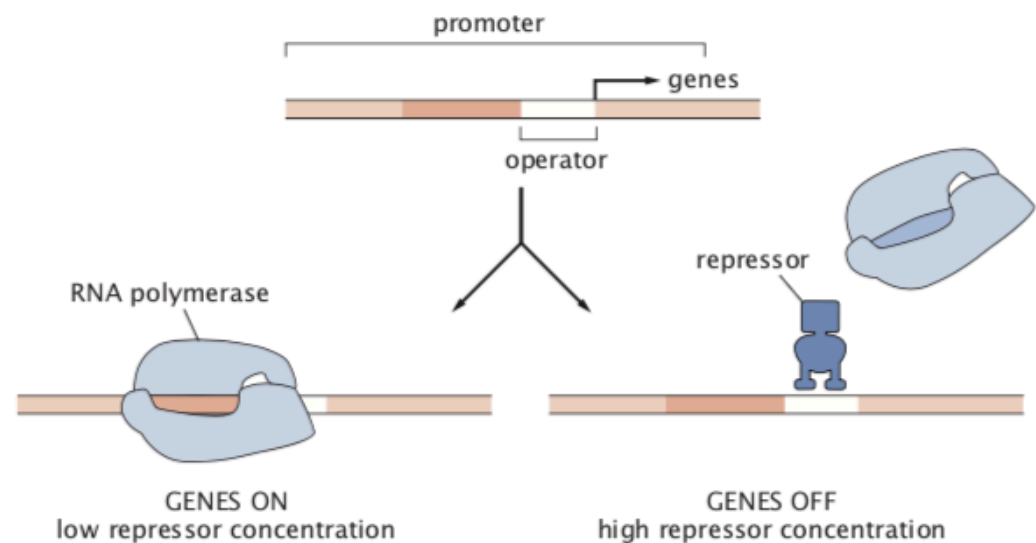
Vamos a extender algunos de los cálculos de hace algunas clases considerando, en primer lugar, a las bacterias. Esto significa que vamos a tratar a la ARN polimerasa como una sola molécula y haremos la pregunta matemática precisa (pero biológicamente simplificada) de si el promotor está ocupado o no por una molécula de ARN polimerasa de este tipo. En el caso eucariota, esta cuestión es menos fácil de plantear, ya que el aparato de transcripción basal consta de muchas partes, todas las cuales deben estar presentes simultáneamente para iniciar la transcripción.

Control negativo

Uno de los mecanismos de control clave de las redes genéticas es la regulación negativa de la transcripción. Esto significa que se regula un paso muy temprano, el que va del ADN a la síntesis del ARNm. Si hay poco o ningún ARNm que codifique una proteína determinada la proteína correspondiente no se producirá.

Este control es ejercido por moléculas proteicas conocidas como represores, e.g., el represor Lac introducido que ya vimos. En el caso de las bacterias, a menudo se puede considerar que los represores llevan a cabo una acción de bloqueo en el sentido de que, a través de interacciones ADN-proteína, ocupan el ADN en una región (llamada operador) que se superpone a la región donde se une la ARN polimerasa (el promotor).

Acá ilustramos la acción de un represor. Se debe notar que la actividad de los represores puede también estar regulada por moléculas pequeñas, o inductores, que pueden unirse y generar un cambio conformacional (o alostérico) que altera la probabilidad de unión del factor de transcripción al ADN.



Imágenes de algunos ejemplos de represores (tomado del libro de Philips et al.).

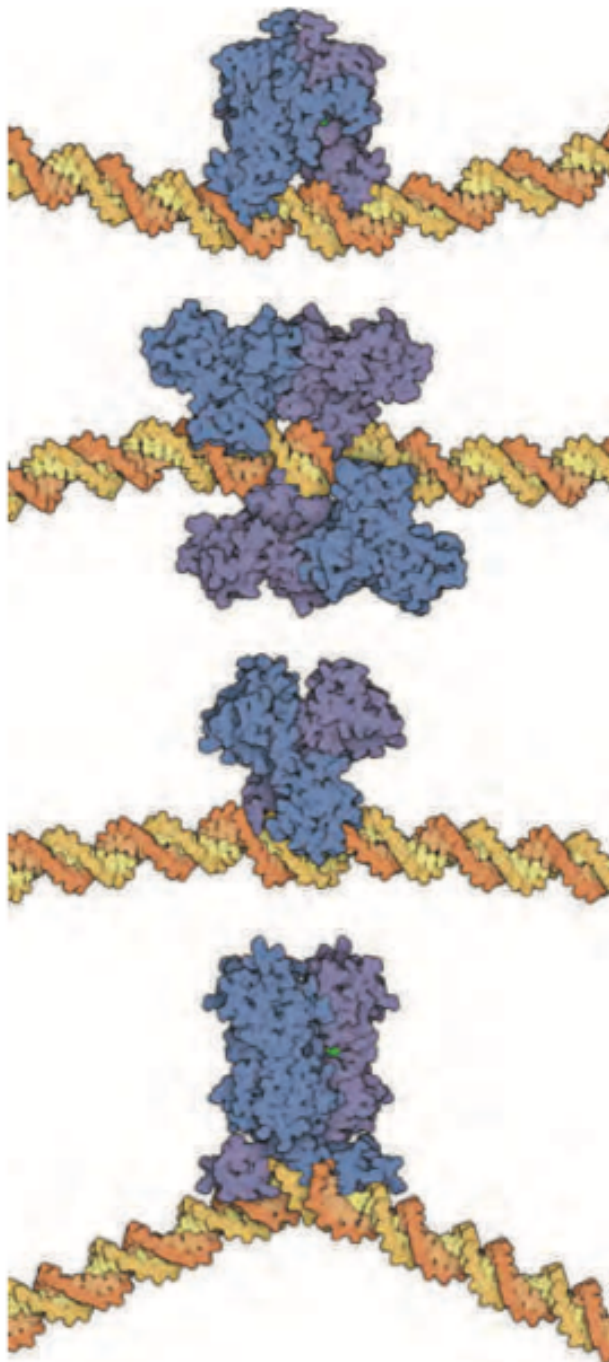


Figure 19.6: Examples of repressor molecules interacting with DNA. From top to bottom, the repressors are TetR (pdb 1QPI), IdeR (pdb 1U8R), FadR (pdb 1HW2), and PurR (pdb 1PNR). The point of the figure is to give an impression of the relative sizes of repressors and their target regions on DNA and to illustrate how these transcription factors deform the DNA double helix in the vicinity of their binding site. These drawings are renditions of actual structures from X-ray crystallography. (Courtesy of D. Goodsell.)

Control positivo

Otro mecanismo clave para alterar el grado en que se expresa un dado gen es la regulación positiva de la transcripción o “reclutamiento regulado”. También en este caso se trata de una regulación temprana que altera la cantidad de ARNm producido y se da a través de una molécula accesoria que aumenta la probabilidad de que la ARN polimerasa se ligue al promotor.

Este mecanismo se basa en la existencia de proteínas distintas de la ARN polimerasa que se unen al ADN y aumentan la probabilidad de que la ARN polimerasa se una al promotor. Así como los represores interfieren con la capacidad de la ARN polimerasa para unirse a su promotor, los activadores se unen en las proximidades del promotor y tienen interacciones con la ARN polimerasa que aumentan su probabilidad de unión al ADN.

El punto clave es que la molécula de ARN polimerasa interactúa no solo con el ADN al que está unida, sino también a través de interacciones “similares a las de ligadura” con el activador.

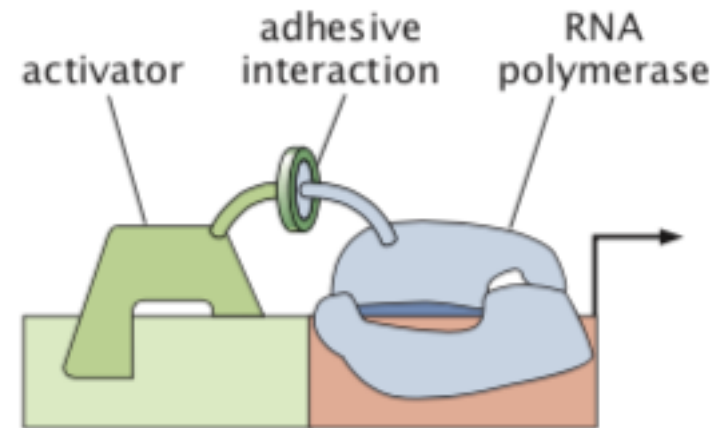


Figure 19.7: The process of activation. Schematic of the way in which activator molecules can recruit the transcription apparatus. Though both the activator and RNA polymerase have their own private interaction energies with the DNA, the enhancement in their occupancies is mediated by the adhesive interaction between them.

Imágenes de algunos ejemplos de activadores (tomado del libro de Philips et al.).

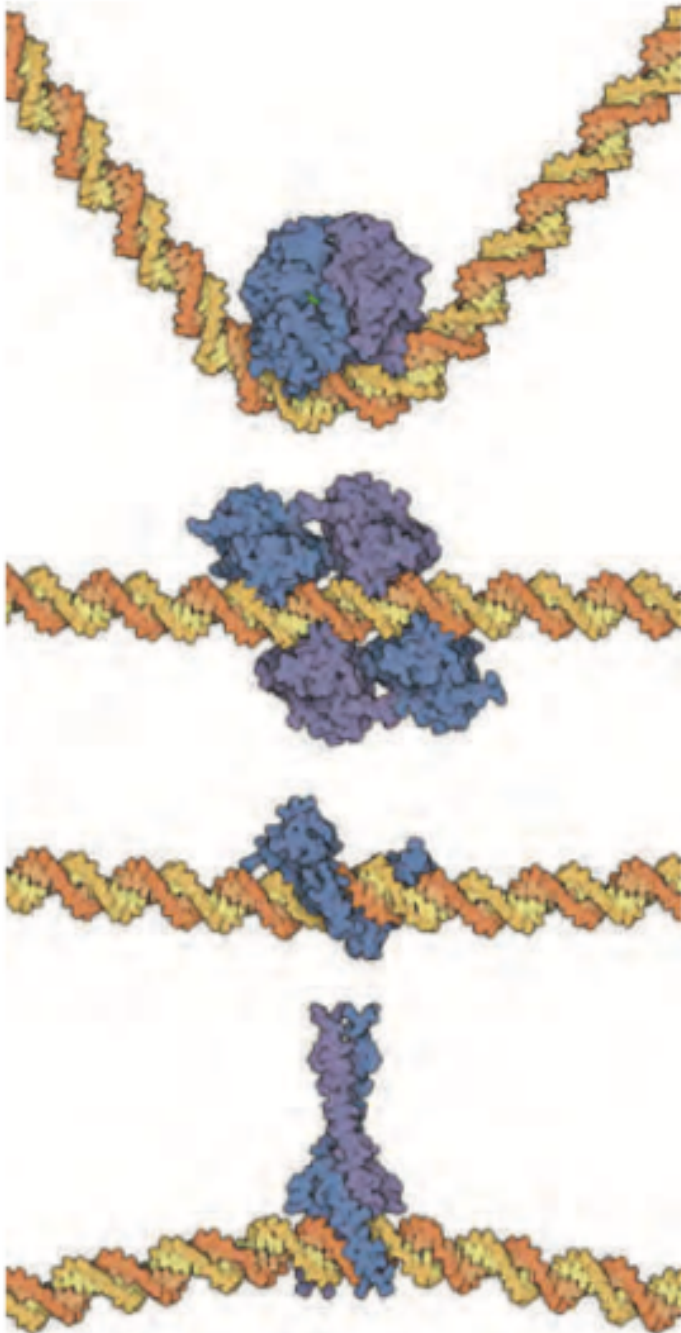


Figure 19.8: Structures of activator molecules. From top to bottom, the activators are CAP (pdb 1CGP), p53 tumor suppressor (pdb 3KMD), zinc finger DNA-binding domain (pdb 2GLI), and leucine zipper DNA-binding domain (pdb 1AN2). (Courtesy of D. Goodsell.)

Una de las ideas generales clave del control transcripcional es la del “reclutamiento molecular” en la que una molécula unida al ADN ayuda a (o impide) que otra molécula se una al ADN para realizar su tarea.

Por ejemplo, la unión de la ARN polimerasa al promotor es ayudada por un factor de transcripción, e.g., la proteína CAP en el caso del operón lac. Desde la perspectiva de las reglas de la mecánica estadística, esto no es más que la conocida idea de la cooperatividad envuelta en diferentes vestimentas verbales.

Los activadores son proteínas que regulan la transcripción uniéndose a un sitio específico en el ADN para reclutar una ARN polimerasa a un sitio promotor cercano. Se ha sugerido que la unión débil e inespecífica de la proteína activadora y la ARN polimerasa puede aumentar en gran medida la probabilidad de que la polimerasa se una al ADN, incluso para las concentraciones muy bajas de proteínas activadoras típicas del entorno celular.

Para evaluar en cuánto difiere esta estrategia de la que tendría lugar si no hubiera moléculas que ayudan, podemos hacer un cálculo de equilibrio en términos de la función de partición como hicimos antes. Lo hacemos para calcular la probabilidad de que la polimerasa se una al promotor en presencia de una proteína activadora utilizando un modelo simple. El punto básico de este cálculo sencillo es mostrar los diferentes estados permitidos en los que se puede encontrar el conjunto formado por la polimerasa y las moléculas activadoras y usar esta enumeración de estados para calcular la probabilidad de que el promotor esté ocupado.

STATE

RENORMALIZED WEIGHT

activator-binding site
promoter

1

$$\frac{P}{N_{NS}} e^{-\beta \Delta \epsilon_{pd}}$$

$$\frac{A}{N_{NS}} e^{-\beta \Delta \epsilon_{ad}}$$

$$\frac{P}{N_{NS}} \frac{A}{N_{NS}} e^{-\beta (\Delta \epsilon_{pd} + \Delta \epsilon_{ad} + \epsilon_{ap})}$$

$$Z(P, A; N_{NS}) = \frac{N_{NS}!}{P! A! (N_{NS} - P - A)!} \times \underbrace{e^{-\beta P \epsilon_{pd}^{NS}}}_{\text{number of arrangements}} \underbrace{e^{-\beta A \epsilon_{ad}^{NS}}}_{\text{weight of each state}}.$$

donde:

En este ejemplo simplificado se ve que hay 4 situaciones posibles (que son las que tendremos en cuenta para calcular la función de partición, Z . Recordemos que Z era el factor de normalización de la probabilidad, por lo que estaba armada como la suma de términos proporcionales a la probabilidad de que el sistema estuviera en cada uno de sus estados permitidos).

Las 4 clases de resultados son: el sitio activador y el promotor están desocupados; solo el promotor está ocupado por la polimerasa; solo el sitio de unión del activador está ocupado por el activador; ambos sitios específicos están ocupados:

$$\begin{aligned} Z_{\text{tot}}(P, A; N_{NS}) = & \underbrace{Z(P, A; N_{NS})}_{\text{empty promoter}} + \underbrace{Z(P-1, A; N_{NS}) e^{-\beta \epsilon_{pd}^S}}_{\text{RNAP}} \\ & + \underbrace{Z(P, A-1; N_{NS}) e^{-\beta \epsilon_{ad}^S}}_{\text{activator}} \\ & + \underbrace{Z(P-1, A-1; N_{NS}) e^{-\beta (\epsilon_{ad}^S + \epsilon_{pd}^S + \epsilon_{pa})}}_{\text{RNAP + activator}}. \end{aligned}$$

es la función de partición para

P moléculas de polimerasa y A de activador que se unen a algunos de los N_{NS} sitios inespecíficos del ADN donde ϵ_{pa} es la energía de la interacción entre el activador y la polimerasa y ϵ_{ad}^S y ϵ_{pd}^{NS} son energías asociadas a la ligadura del activ. a sitios específicos y no específicos.

Suponiendo que: $N_{NS} \gg A + P$ se puede aproximar: $N_{NS}!/(N_{NS} - A - P)! \approx (N_{NS})^{A+P}$

Para calcular la probabilidad de que la RNA polimerasa esté ligada al promotor tenemos en cuenta los 2 casos en que esto ocurre (con o sin el activador ligado al sitio específico):

$$p_{\text{bound}}(P, A; N_{NS}) = \frac{Z(P-1, A; N_{NS})e^{-\beta\epsilon_{pd}^S} + Z(P-1, A-1; N_{NS})e^{-\beta(\epsilon_{ad}^S + \epsilon_{pd}^S + \epsilon_{pa})}}{Z_{\text{tot}}(P, A; N_{NS})}$$

que se puede reescribir como:

$$p_{\text{bound}}(P, A; N_{NS}) = \frac{1}{1 + [N_{NS}/PF_{\text{reg}}(A)]e^{\beta\Delta\epsilon_{pd}}},$$

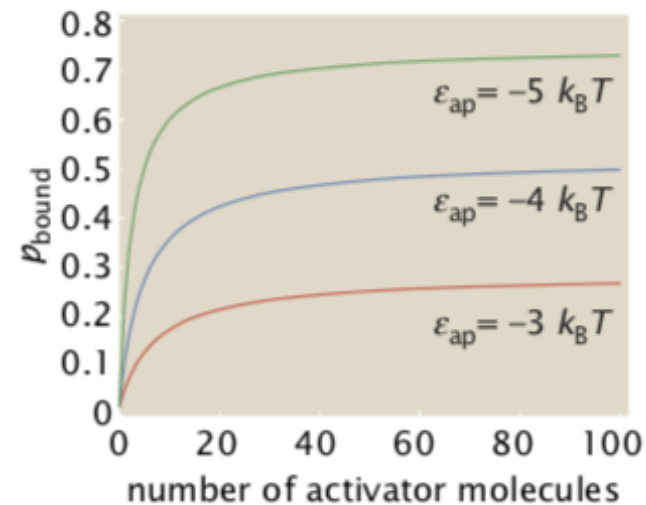
con:

$$F_{\text{reg}}(A) = \frac{1 + (A/N_{NS})e^{-\beta\Delta\epsilon_{ad}}e^{-\beta\epsilon_{ap}}}{1 + (A/N_{NS})e^{-\beta\Delta\epsilon_{ad}}},$$

donde: $\Delta\epsilon_{pd} = \epsilon_{pd}^S - \epsilon_{pd}^{NS}$ y $\Delta\epsilon_{ad} = \epsilon_{ad}^S - \epsilon_{ad}^{NS}$.

F_{reg} es tal que, en el límite en el que la interacción de adhesión entre la polimerasa y el activador tiende a cero, F_{reg} tiende a 1. Por otro lado, para valores negativos de esta interacción (es decir, cuando al activador y a la polimerasa les "gusta estar cerca uno del otro"), el factor de regulación es > 1 . Esto último se traduce en un aumento efectivo en el número de moléculas de polimerasa ligadas.

En la figura se muestra la probabilidad de unión de la RNA polimerasa como función del número de moléculas de activador, A , para $P = 500$, $\Delta\epsilon_{pd} = -5,3 k_B T$ y $\Delta\epsilon_{ad} = -13,12 k_B T$. Las 3 curvas corresponden a distintas energías de interacción entre la polimerasa y el activador.



Un simple cambio en el número efectivo de moléculas de ARN polimerasa ($P \rightarrow P_{\text{eff}}$) es suficiente para capturar la acción de activadores y represores. Esta interpretación del significado del factor de regulación se muestra en la Figura.

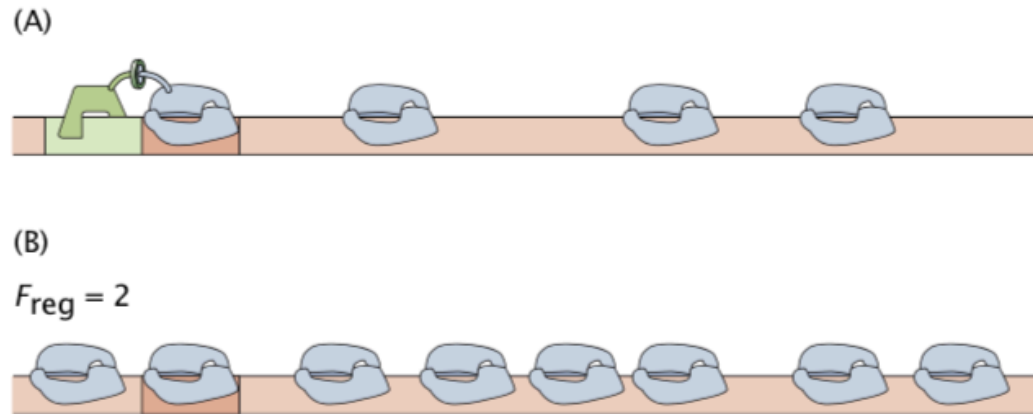


Figure 19.11: Regulation factor and the effective number of polymerase molecules. The presence of activators is equivalent to a problem with just polymerase molecules but a larger number of them. (A) The “bare” problem with activators and polymerase present. (B) The “effective” problem in which the presence of activators is treated as a change in the number of polymerase molecules.

La presencia de activadores es equivalente a que el número de moléculas de ARN polimerasa cambie de P a $F_{\text{reg}}P$. Para el caso de los activadores, el factor de regulación es mayor que 1 y conduce a un aumento efectivo del número de moléculas de polimerasa.

Para comparar los casos con y sin activadores calculamos:

$$\text{fold-change} = \frac{p_{\text{bound}}(A \neq 0)}{p_{\text{bound}}(A = 0)} = \frac{1 + (N_{\text{NS}}/P)e^{\beta\Delta\epsilon_{\text{pd}}}}{1 + [N_{\text{NS}}/PF_{\text{reg}}(A)]e^{\beta\Delta\epsilon_{\text{pd}}}}.$$

Acá se ve que si un promotor es débil se satisface la condición $(N_{\text{NS}}/P)e^{\beta\Delta\epsilon_{\text{pd}}} \gg 1$, lo que implica que el fold-change se puede reescribir como $\text{fold-change} \approx F_{\text{reg}}(A)$, donde supusimos que $(N_{\text{NS}}/PF_{\text{reg}})e^{\beta\Delta\epsilon_{\text{pd}}} \gg 1$, es decir, que no es fuerte incluso en el caso regulado.

La conclusión es que, en el caso de un promotor débil, los detalles vinculados al promotor, como su energía de ligadura, desaparecen del cálculo.

Con

$$\text{fold-change} = \frac{p_{\text{bound}}(A \neq 0)}{p_{\text{bound}}(A = 0)} = \frac{1 + (N_{\text{NS}}/P)e^{\beta\Delta\epsilon_{\text{pd}}}}{1 + [N_{\text{NS}}/PF_{\text{reg}}(A)]e^{\beta\Delta\epsilon_{\text{pd}}}}.$$

obtenemos una figura como ésta:

La fórmula muestra también que si un promotor es suficientemente débil de modo que $(N_{\text{NS}}/P)e^{\beta\Delta\epsilon_{\text{pd}}} \gg 1$ el fold-change se puede reescribir como $\text{fold-change} \approx F_{\text{reg}}(A)$. donde supusimos que $(N_{\text{NS}}/PF_{\text{reg}})e^{\beta\Delta\epsilon_{\text{pd}}} \gg 1$, es decir, que no es fuerte tampoco en el caso regulado.

La conclusión es que, en el caso de un promotor débil, los detalles vinculados al promotor, como su energía de ligadura, desaparecen del cálculo.

La lógica seguida para calcular el cambio en la probabilidad de ligadura en presencia de un activador puede usarse para analizar el problema de la represión.

De nuevo tenemos considerar las distintas situaciones posibles para calcular la función de partición.

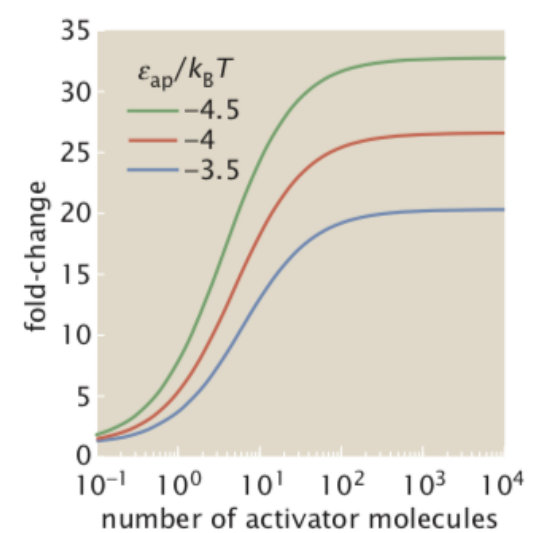
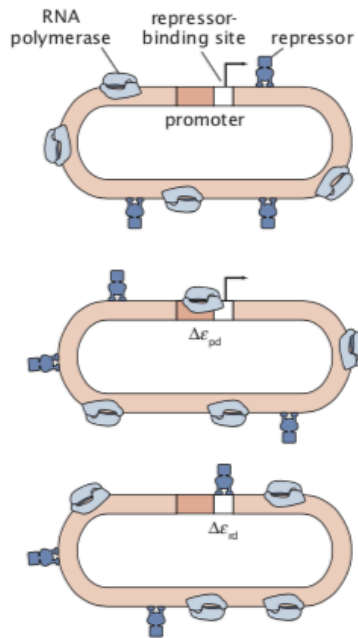


Figure 19.12: Fold-change due to activators. Fold-change in gene expression as a function of the number of activators for different activator–RNA polymerase interaction energies using $P = 500$, $\Delta\epsilon_{\text{pd}} = -5.3 k_B T$, and $\Delta\epsilon_{\text{ad}} = -13.12 k_B T$ based on *in vitro* measurements.



1

$$\frac{P}{N_{NS}} e^{-\beta \Delta \epsilon_{pd}}$$

$$\frac{R}{N_{NS}} e^{-\beta \Delta \epsilon_{rd}}$$

$$Z(P, R; N_{NS}) = \frac{N_{NS}!}{P!R!(N_{NS} - P - R)!} e^{-\beta P \epsilon_{pd}^{NS}} e^{-\beta R \epsilon_{rd}^{NS}},$$

donde: ϵ_{rd}^{NS} describe la unión inespecífica del represor al ADN (ϵ^S será la energía de unión específica del represor a su operador).

Para escribir la función de partición total sumamos todos los estados en los que el promotor está ocupado por una molécula represora o por una molécula de ARN polimerasa. Este modelo no incluye detalles como como si la ARN polimerasa puede estar en su promotor al mismo tiempo que el represor está unido a su operador o no:

$$Z_{\text{tot}}(P, R; N_{NS}) = \underbrace{Z(P, R; N_{NS})}_{\text{empty promoter}} + \underbrace{Z(P - 1, R; N_{NS}) e^{-\beta \epsilon_{pd}^S}}_{\text{RNAP on promoter}}$$

$$+ \underbrace{Z(P, R - 1; N_{NS}) e^{-\beta \epsilon_{rd}^S}}_{\text{repressor on promoter}}.$$

La probabilidad de que el promotor esté ocupado por la ARN polimerasa está dada entonces por:

$$p_{\text{bound}}(P, R; N_{NS})$$

$$= \frac{Z(P - 1, R; N_{NS}) e^{-\beta \epsilon_{pd}^S}}{Z(P, R; N_{NS}) + Z(P - 1, R; N_{NS}) e^{-\beta \epsilon_{pd}^S} + Z(P, R - 1; N_{NS}) e^{-\beta \epsilon_{rd}^S}}.$$

Que puede reescribirse como:

$$p_{\text{bound}}(P, R; N_{NS}) = \frac{1}{1 + (N_{NS}/P) e^{\beta(\epsilon_{pd}^S - \epsilon_{pd}^{NS})} [1 + (R/N_{NS}) e^{-\beta(\epsilon_{rd}^S - \epsilon_{rd}^{NS})}]}.$$

La expresión $p_{\text{bound}}(P, R; N_{\text{NS}}) = \frac{1}{1 + (N_{\text{NS}}/P)e^{\beta(\epsilon_{\text{pd}}^S - \epsilon_{\text{pd}}^{\text{NS}})}[1 + (R/N_{\text{NS}})e^{-\beta(\epsilon_{\text{rd}}^S - \epsilon_{\text{rd}}^{\text{NS}})}]}$ puede ponerse en forma más compacta usando: $F_{\text{reg}}(R) = \left(1 + \frac{R}{N_{\text{NS}}}e^{-\beta\Delta\epsilon_{\text{rd}}}\right)^{-1}$, donde: $\Delta\epsilon_{\text{rd}} = \epsilon_{\text{rd}}^S - \epsilon_{\text{rd}}^{\text{NS}}$

En el caso de la represión, el factor de regulación es tal que: $F_{\text{reg}} < 1$, lo que es matemáticamente equivalente a suponer como una reducción en el número efectivo de moléculas de ARN polimerasa.

El libro de R. Philips tiene toda una parte de exploración numérica en las pp 817-819 que, quienes estén interesados, pueden hacer. También ilustra la represión con este ejemplo:

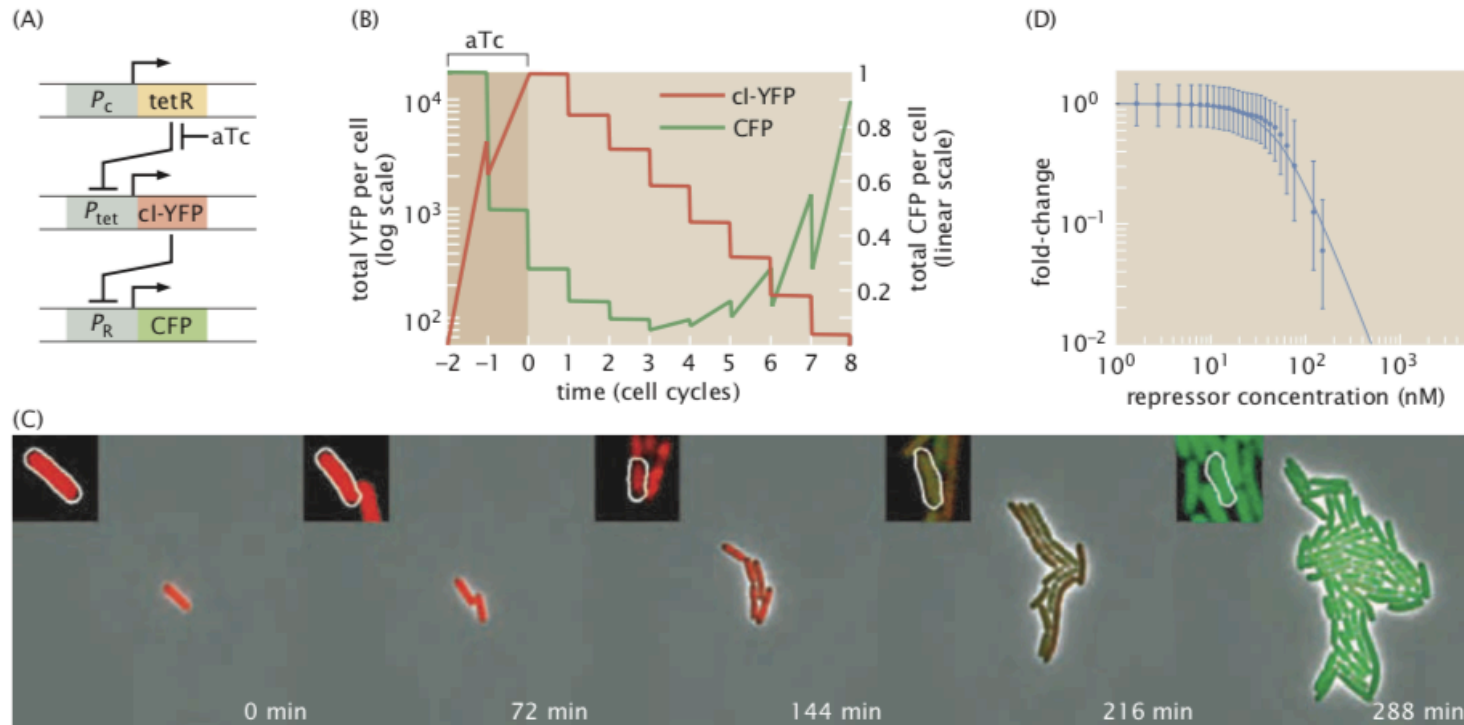


Figure 19.15: Dilution experiment and the measurement of fold-change in repression. (A) Diagram of the circuit. In the absence of the inducer aTc, the repressor TetR shuts down production of the transcription factor cl fused to YFP. This transcription factor, in turn, regulates the expression of the reporter CFP. (B) Schematic of the time course of an experiment. Adding aTc for a short period of time leads to the production of cl-YFP . Upon removal of aTc, no new cl-YFP is produced. As a result, in each new generation, there will be decreasing numbers of cl-YFP per cell, resulting in an ever-higher rate of expression of the downstream CFP gene. This dilution also permits the calibration of YFP fluorescence into absolute numbers of cl-YFP as discussed in the text. (C) Representative snapshots from the time course of an experiment. (D) Fold change (1/repression) as a function of cl repressor concentration measured using the dilution method. (Adapted from N. Rosenfeld et al. *Science* 307: 1962, 2005.)

La regulación real en realidad involucra la acción simultánea de la activación y la represión.

Si, por ejemplo, consideramos estas situaciones

STATE	RENORMALIZED WEIGHT
	1
	$\frac{P}{N_{NS}} e^{-\beta \Delta \epsilon_{pd}}$
	$\frac{A}{N_{NS}} e^{-\beta \Delta \epsilon_{ad}}$
	$\frac{P}{N_{NS}} \frac{A}{N_{NS}} e^{-\beta (\Delta \epsilon_{pd} + \Delta \epsilon_{ad} + \epsilon_{ap})}$
	$\frac{R}{N_{NS}} e^{-\beta \Delta \epsilon_{rd}}$
	$\frac{A}{N_{NS}} \frac{R}{N_{NS}} e^{-\beta (\Delta \epsilon_{ad} + \Delta \epsilon_{rd})}$

La función de partición resulta:

$$\begin{aligned}
 Z_{\text{tot}}(P, A, R; N_{NS}) &= \underbrace{Z(P, A, R; N_{NS})}_{\text{empty promoter}} + \underbrace{Z(P-1, A, R; N_{NS}) e^{-\beta \epsilon_{pd}^S}}_{\text{RNAP}} \\
 &+ \underbrace{Z(P, A-1, R; N_{NS}) e^{-\beta \epsilon_{ad}^S}}_{\text{activator}} + \underbrace{Z(P-1, A-1, R; N_{NS}) e^{-\beta (\epsilon_{ad}^S + \epsilon_{pd}^S + \epsilon_{pa})}}_{\text{RNAP + activator}} \\
 &+ \underbrace{Z(P, A, R-1; N_{NS}) e^{-\beta \epsilon_{rd}^S}}_{\text{repressor}} + \underbrace{Z(P, A-1, R-1; N_{NS}) e^{-\beta (\epsilon_{ad}^S + \epsilon_{rd}^S)}}_{\text{activator + repressor}}.
 \end{aligned}$$

de donde:

$$\begin{aligned}
 p_{\text{bound}}(P, A, R; N_{NS}) &= \frac{Z(P-1, A, R; N_{NS}) e^{-\beta \epsilon_{pd}^S} + Z(P-1, A-1, R; N_{NS}) e^{-\beta (\epsilon_{ad}^S + \epsilon_{pd}^S + \epsilon_{pa})}}{Z_{\text{tot}}(P, A, R; N_{NS})}.
 \end{aligned}$$

que puede escribirse como:

$$p_{\text{bound}}(P, A, R; N_{NS}) = \frac{1}{1 + [N_{NS}/PF_{\text{reg}}(A, R)] e^{\beta (\epsilon_{pd}^S - \epsilon_{pd}^{NS})}},$$

Con:

$$\begin{aligned}
 F_{\text{reg}}(A, R) &= \frac{1}{1 + (A/N_{NS}) e^{-\beta (\Delta \epsilon_{ad} + \epsilon_{ap})}} \\
 &= \frac{1}{1 + (A/N_{NS}) e^{-\beta \Delta \epsilon_{ad}} + (R/N_{NS}) e^{-\beta \Delta \epsilon_{rd}} + (A/N_{NS})(R/N_{NS}) e^{-\beta (\Delta \epsilon_{ad} + \Delta \epsilon_{pd})}}.
 \end{aligned}$$

Se muestra en la figura el fold-change de la expresión génica debido a esta combinación cuando el promotor es débil. Esta figura corresponde a la situación que se da en el operón lac. Aquí, en lugar de variar el número intracelular de factores de transcripción, se adopta el enfoque más simple de medir la actividad del promotor lac en función de los dos inductores que controlan la unión del represor y el activador al ADN (IPTG y cAMP, respectivamente).

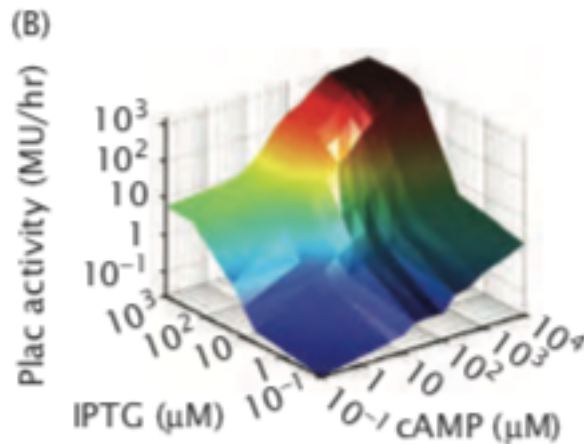
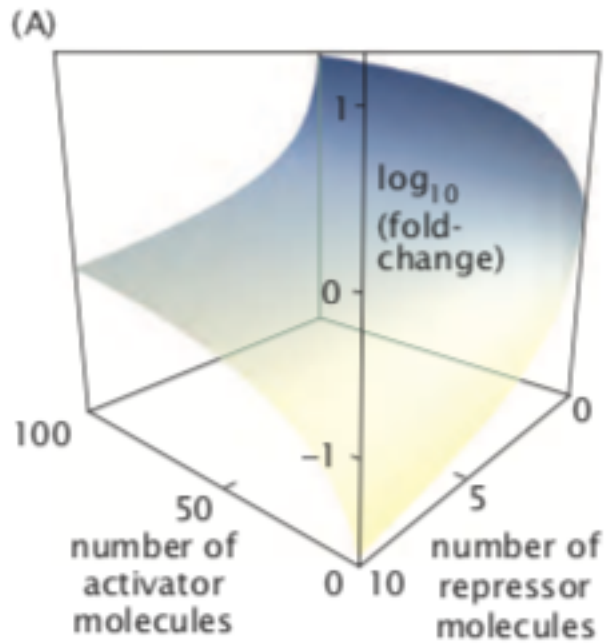


Figure 19.19: Combined regulation by repressor and activator. (A) The fold-change in gene expression as a function of the number of transcription factors shows their combinatorial action. The parameters used are $\Delta\epsilon_{ad} = -10 k_B T$, $\epsilon_{ap} = -3.9 k_B T$ and $\Delta\epsilon_{rd} = -16.9 k_B T$. (B) Activity of the *lac* operon measured in Miller units (MU) per hour as a function of the concentration of IPTG and cAMP, which regulate the binding of Lac repressor and CRP to the DNA, respectively. (B, adapted from T. Kuhlman et al., *Proc. Natl. Acad. Sci. USA* 104:6043, 2007.)

Hasta ahora, discutimos los casos de represión simple, activación simple y loop de ADN. En cada uno de estos casos el punto de contacto experimental fue el fold-change, que dice cómo cambia la expresión en presencia de diferentes factores reguladores. Esta figura resume cómo los modelos termodinámicos descritos se pueden usar para predecir la respuesta de entrada-salida de estas arquitecturas reguladoras.

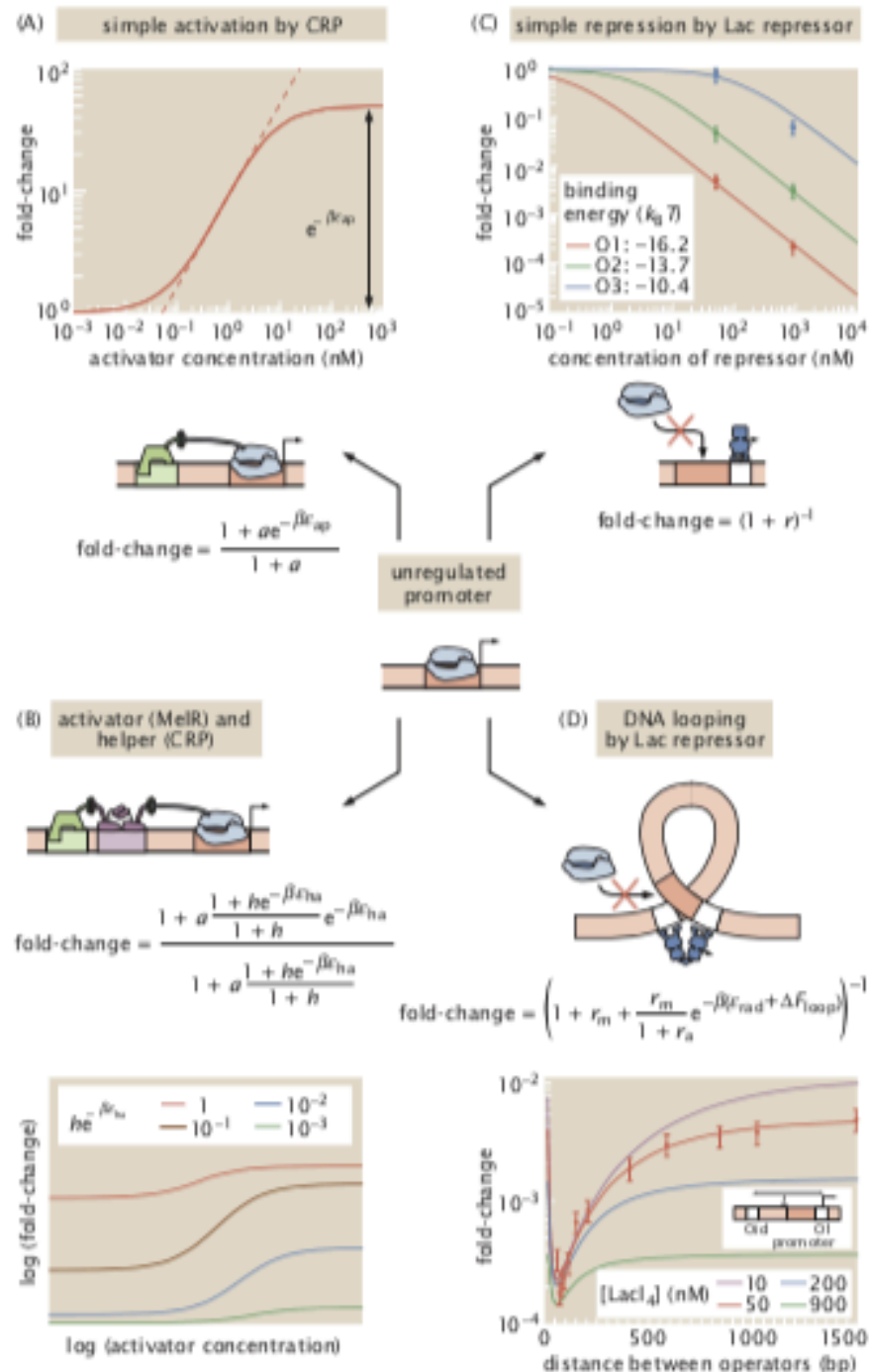


Figure 19.29: Thermodynamic models for diverse regulatory architectures. The thermodynamic models result in a predicted fold-change as a function of parameters such as operator binding strengths and transcription factor copy numbers. The resulting fold-change function serves as a governing equation dictating the regulatory output. The lowercase variables (a, h, r , etc.) correspond to $x = (X/N_{NS})e^{-\beta \Delta \epsilon_{x,d}}$, where X is the intracellular number of the particular transcription factor, $\Delta \epsilon_{x,d}$ its interaction energy with the DNA, and $N_{NS} = 5 \times 10^6$. (A) In simple activation, an activator recruits RNA polymerase to the promoter by interacting with it ($\Delta \epsilon_{ad} = -13.8 \text{ } k_B T$, $\epsilon_{ap} = 3.9 \text{ } k_B T$). (B) A helper molecule can recruit an activator to the promoter, which in turn can recruit RNA polymerase. (C) In simple repression, a repressor binds to a site overlapping the promoter, which results in the exclusion of RNA polymerase from that site. (D) Some repressors can also bind to multiple sites simultaneously, which results in an increase of the level of repression. The looping probability will depend on physical characteristics of the loop, such as its length ($\Delta \epsilon_{end} = -16.7 \text{ } k_B T$, $\Delta \epsilon_{rad} = -18.4 \text{ } k_B T$, and $\Delta F_{loop} = A/L + B \ln L - CL - E$, with $A = 140.6 \text{ } k_B T \times \text{bp}$, $B = 2.52 \text{ } k_B T$, $C = 1.4 \times 10^{-3} \text{ } k_B T/\text{bp}$ and $E = 19.9 \text{ } k_B T$). We assume that one molecule per cell corresponds to 1 nM. (Adapted from L. Bintu et al., *Curr. Opin. Genet. Dev.* 15:125, 2005.)

Conceptualmente, lo que se busca son descripciones de estas redes en las que se puedan identificar las diversas “perillas” reguladoras como las que se muestran en la figura de esta transparencia (Fig. 19.30) y ajustar estas perillas tanto teóricamente como en el contexto de experimentos cuantitativos.

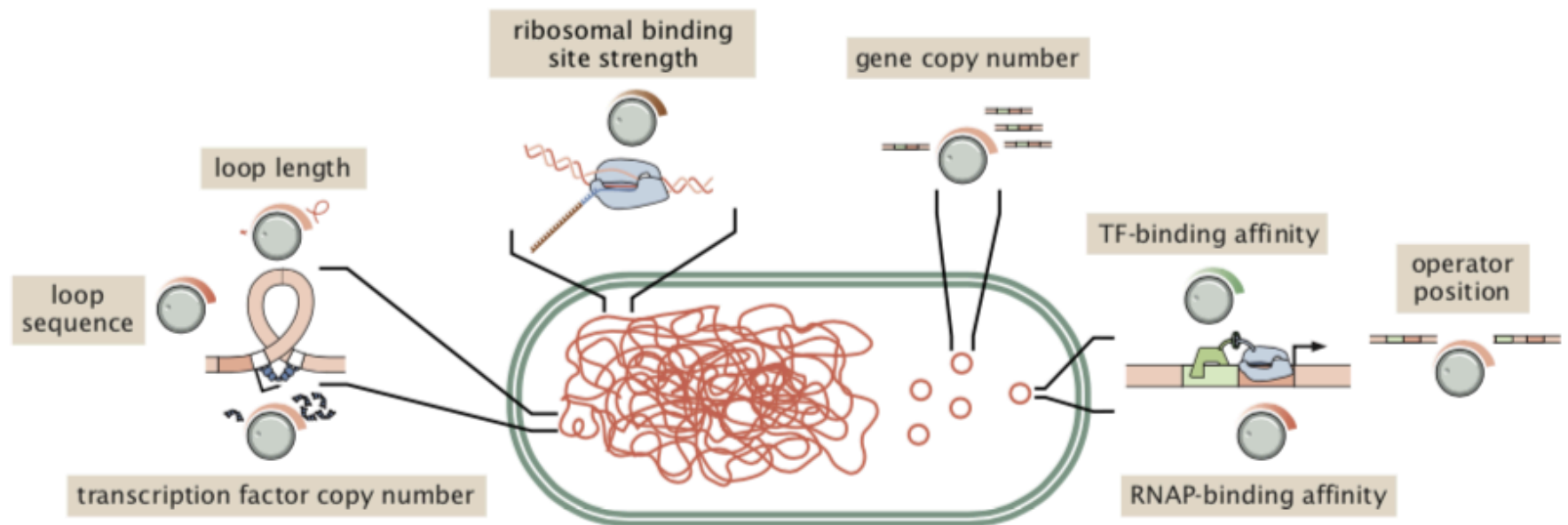


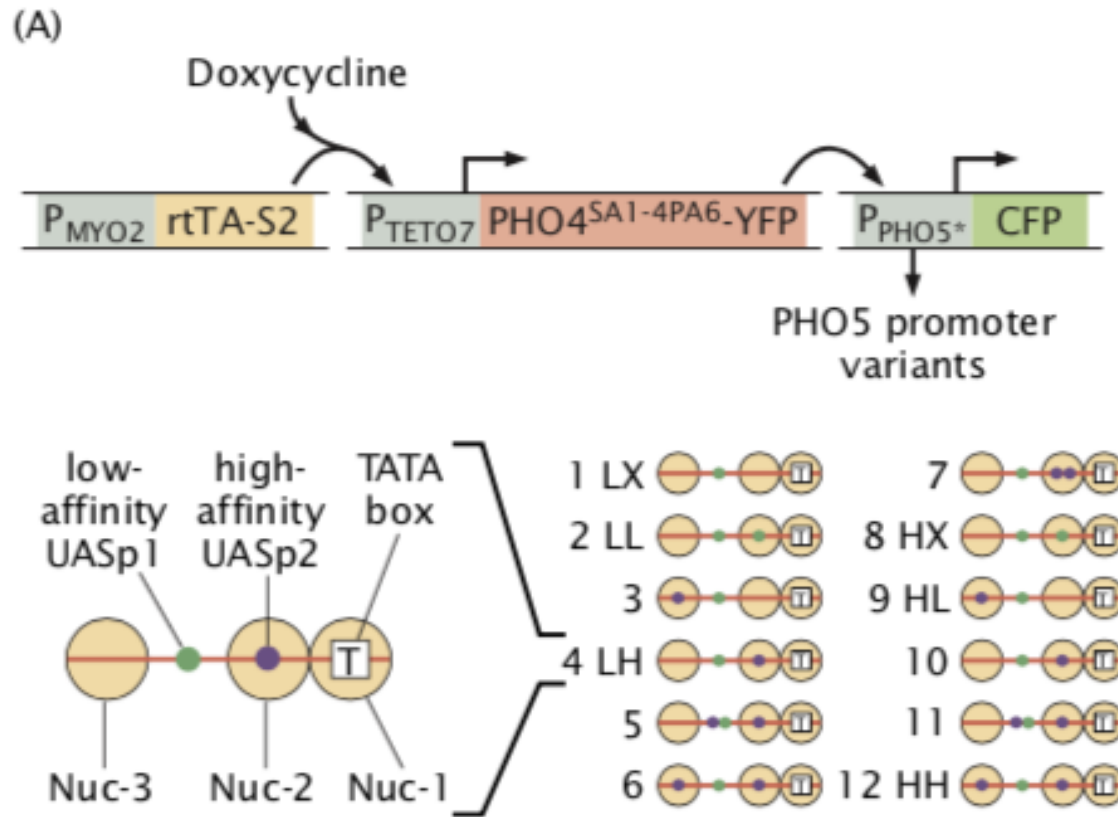
Figure 19.30: Dialing-in transcriptional output. Many parameters can be used to quantitatively tune the level of gene expression, including the number of copies of genes and the transcription factors that control them, as well as the binding strengths of a suite of different proteins that interact with the DNA during gene expression.

Las redes en células eucariotas se pueden estudiar de forma similar pero, como siempre, todo es más complicado que en bacterias.

En particular, una gran cantidad de especies moleculares contribuyen colaborativamente impulsar la transcripción. La misma descripción termodinámica aplicada hasta ahora se ha utilizado para pensar en cuestiones como la ocupación de nucleosomas y cómo ésta interfiere con la transcripción para una serie de diferentes arquitecturas reguladoras eucariotas.

Se sabe que los nucleosomas están posicionados sobre los genomas eucariotas. Se los encuentra en regiones reguladoras, lo que hace que los sitios de unión al factor de transcripción estén ocluidos. Uno de los sistemas en los que se ha explorado cuantitativamente la interacción entre la ocupación nucleosómica y el nivel de expresión génica es el promotor PHO5 en la levadura.

El promotor PHO5 es activado por PHO4, que se une a dos sitios, uno dentro de un nucleosoma y el otro no. Además, hay una caja de unión TATA en el nucleosoma -1. Este es el sitio de unión para la proteína de unión a la caja TATA (TBP), que es fundamental para la activación del gen.



Experimentalmente se puede mutar la región reguladora para mover los sitios de unión a los nucleosomas o fuera de ellos y cambiar sus afinidades. El posicionamiento de estos sitios de unión con respecto a los de los nucleosomas es importante. En particular, al comparar arquitecturas con sitios de idéntica afinidad, pero diferente oclusión por nucleosomas, las características de la función de entrada-salida cambian apreciablemente.

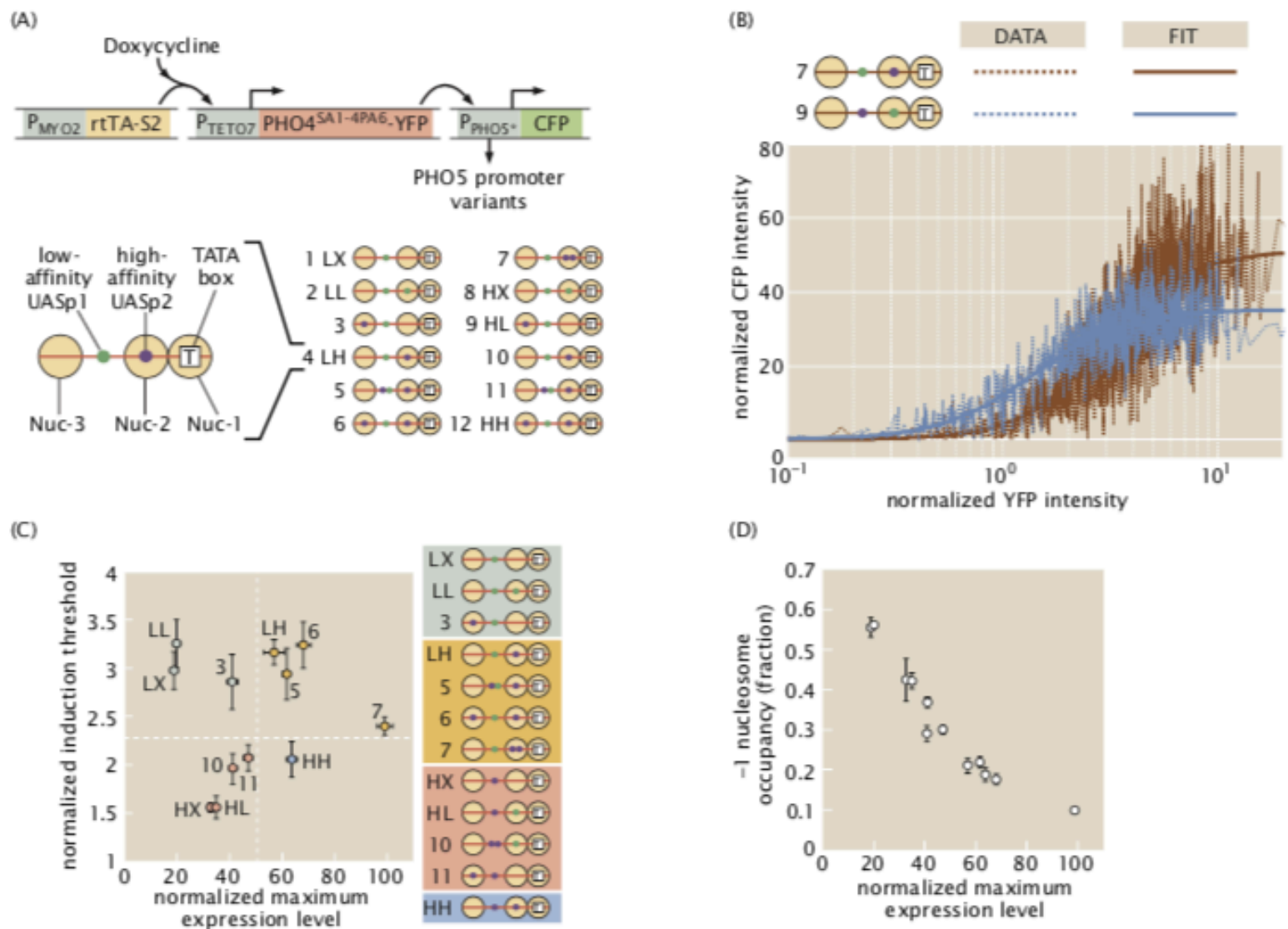


Figure 19.31: Role of nucleosomes in transcriptional regulation. (A) A PHO5 promoter expressing CFP is regulated by PHO4-YFP, with two target binding sites, UASp1 and UASp2. This promoter has a TATA box, which is bound by the TBP activator. The position of nucleosomes in this regulatory region has been mapped such that mutants can be generated where the various promoter features are occluded or not by nucleosomes. (B) Two architectures having the same binding sites but different occlusion geometries by nucleosomes show markedly different input-output functions. (C) Several variants for the PHO5 promoter result in input-output functions with drastically different induction thresholds and maximum levels of expression. (D) The maximum level of expression shows a strong correlation with the occupancy of nucleosome -1, which occludes the TATA box. (Adapted from Kim and O'Shea, *Nat. Struct. Mol. Biol.* 15:1192, 2008.)

Otro desafío para descifrar la regulación transcripcional eucariota se centra en estudiar cómo varía la transcripción en el espacio y en el tiempo durante el desarrollo en organismos multicelulares. El organismo más estudiado es uno que ya vimos: la mosca de la fruta *Drosophila melanogaster*.

Ya mencionamos que durante las etapas iniciales de desarrollo el embrión de mosca expresa una batería de factores de transcripción en una cascada que define dominios de expresión cada vez más nítidos.

Una de las arquitecturas transcripcionales que se ha estudiado con más detalle está relacionada con la activación del factor de transcripción Hunchback por parte del factor de transcripción, Bicoid.

Ya contamos que Bicoid se expresa en un perfil exponencial a lo largo del eje anteroposterior del embrión en desarrollo. La activación por Bicoid se realiza mediante la unión a seis sitios de distintas afinidades que se encuentran aguas arriba del promotor Hunchback. El patrón resultante de expresión de Hunchback presenta un dominio con un límite en aproximadamente el 50% de la longitud del embrión. Esto se ve en la figura que sigue.

Bicoid y Hunchback en embriones de mosca.

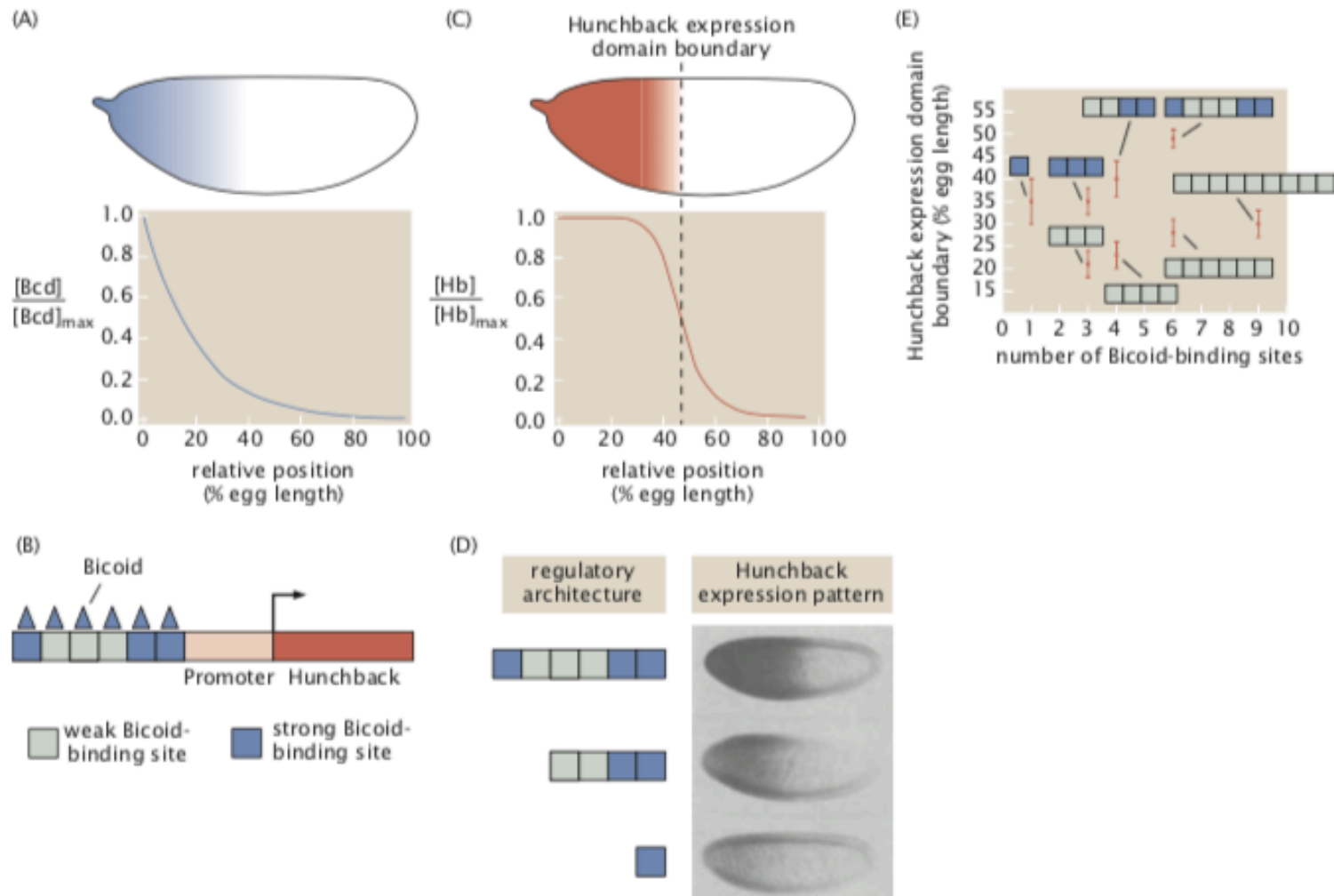


Figure 19.32: Systematic analysis of gene expression in *Drosophila*. (A) The Bicoid transcription factor is expressed in an exponential profile from the anterior to the posterior end of the fly embryo. (B) Bicoid acts as an activator of the Hunchback transcription factor by binding to six binding sites of different strengths located upstream from the Hunchback promoter. (C) The resulting pattern of Bicoid-dependent Hunchback expression domain presents a sharp boundary at about 50% of the embryo length. (D) By creating constructs with different numbers and affinities of binding sites, the boundary of the expression domain can be shifted systematically. (E) Hunchback domain boundary position for several regulatory architectures. (D, E, adapted from W. Driever et al., *Nature* 340:363, 1989.)

El cálculo termodinámico aplicado a Bicoid y Hunchback.

La idea es calcular la cantidad de gen hunchback expresado en una posición específica a lo largo del eje anteroposterior del embrión, dada la cantidad de Bicoid en esa misma posición.

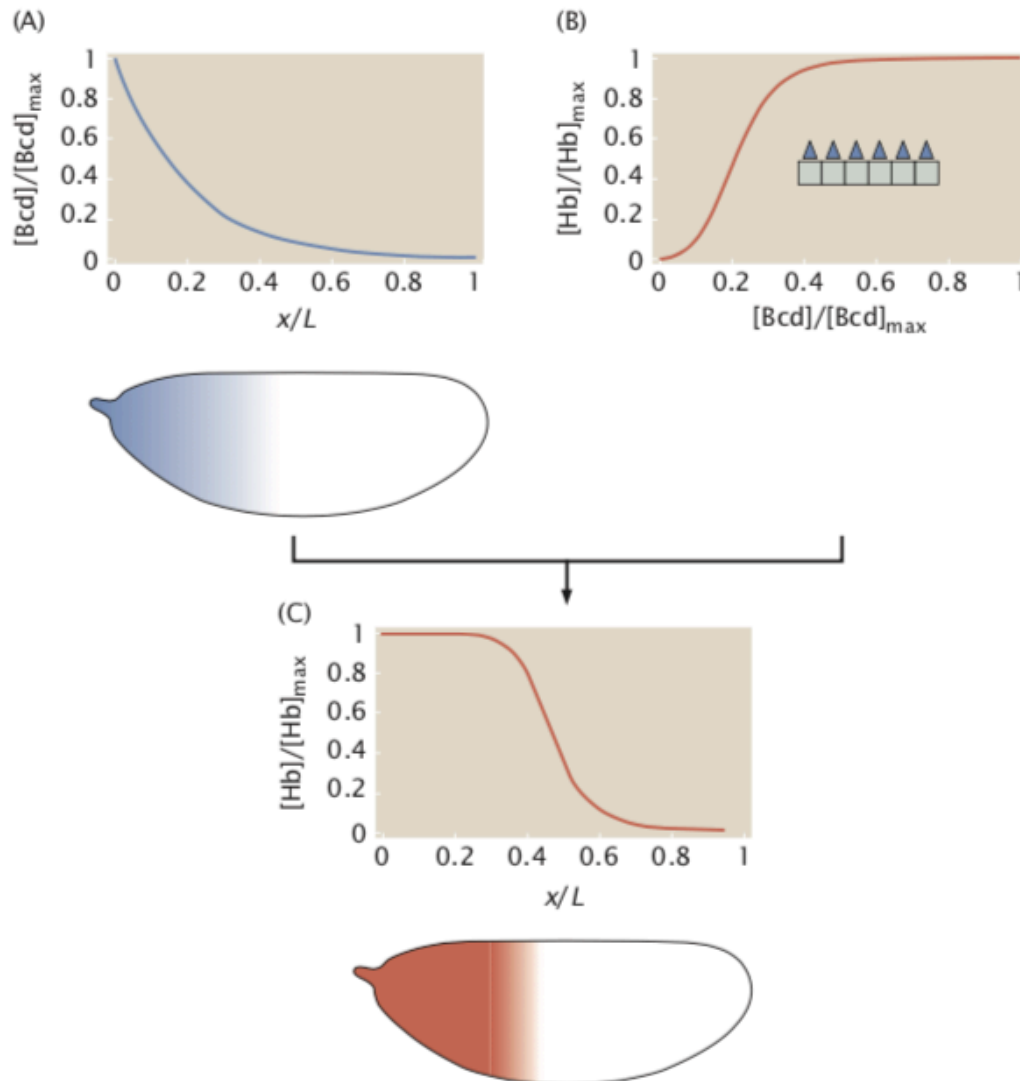


Figure 20.7: Establishment of the Hunchback profile. (A) Bicoid gradient along the anterior-posterior axis. (B) Amount of *hunchback* gene expressed as a function of the concentration of Bicoid. $[Hb]/[Hb]_{\max}$ is identified with p_{on} , which is given by Equation 20.17, where we have assumed $\varepsilon_{\text{on}} - \varepsilon_{\text{off}} = 5k_B T$, and K_d is chosen to correspond to the concentration of Bicoid at $x/L = 0.5$, that is, $K_d = 0.08[Bcd]_{\max}$. (Experiments find that the Hunchback step occurs at roughly the midpoint between the anterior and the posterior end of the embryo.) The inset shows the six known Bicoid-binding sites responsible for Hunchback expression at this stage of development. (C) Concentration of Hunchback protein along the anterior-posterior axis is obtained by using the Bicoid gradient, Equation 20.18, as input into Equation 20.17.

Se cree que Bicoid controla la transcripción de *hunchback* uniéndose a seis sitios de unión distintos en la región del ADN. Esto se puede describir usando un modelo tipo Monod-Wyman-Changeux (MWC). La idea del modelo es que, como resultado de la compactación espacial del gen *hunchback*, el sistema puede estar en un estado transcripcionalmente activo ("on") o en un estado compactado ("off"). Esta compactación espacial podría deberse, e.g., a la presencia de nucleosomas que ocluyen la unión de factores de transcripción al ADN.

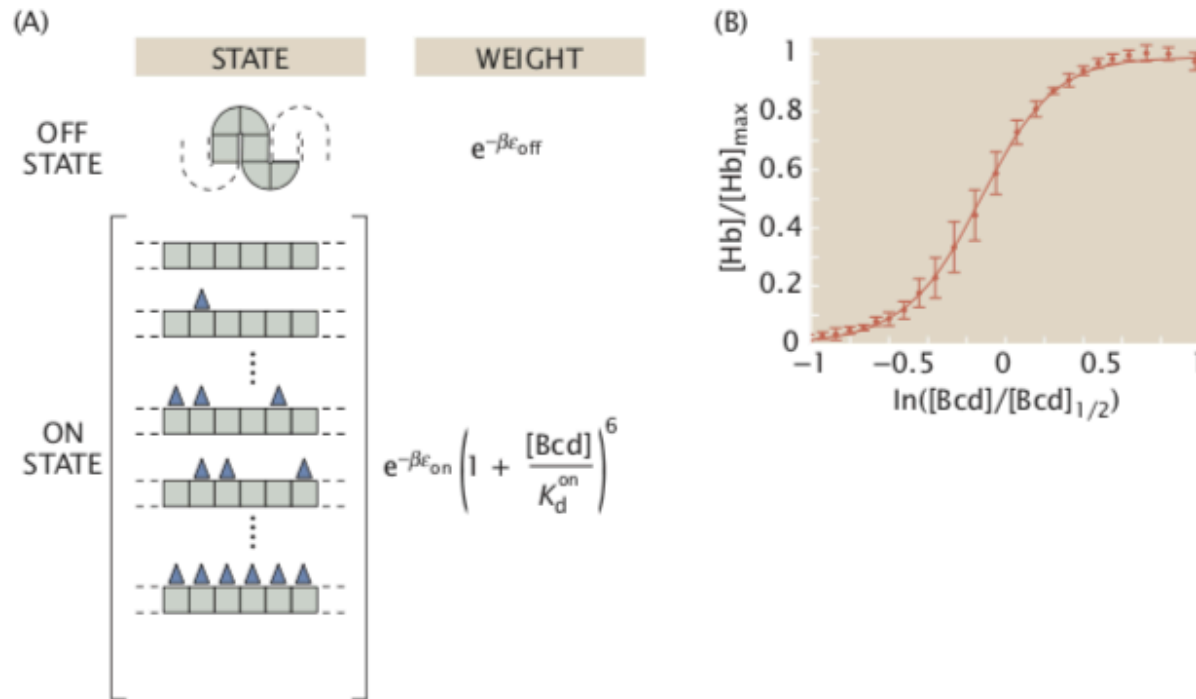


Figure 20.8: Bicoid activation of *hunchback*. (A) States and weights for the MWC model of regulation of *hunchback* expression by Bicoid. The DNA is imagined to have two states, one available for Bicoid binding and the other for which such binding is forbidden. There are six binding sites for Bicoid within the binding region. (B) Mean output level of Hunchback as a function of the input level of Bicoid averaged over nine embryos. (B, adapted from T. Gregor et al., *Cell* 130:153, 2007.)

Para que Bicoid actúe como activador transcripcional, debe favorecer el estado "on". Suponemos que en este estado los 6 sitios de unión de Bicoid están disponibles y pueden estar o no ocupados por moléculas de Bicoid. En el estado "off" los sitios de unión no están disponibles. Esto podría suceder si el ADN que contiene los sitios de unión se enrollara en los nucleosomas cuando el gen está en el estado "off", ocluyendo así estos sitios de unión.

Calculamos entonces la probabilidad de que el sistema esté en el estado on (p_{on}) y la equiparamos a la cantidad de Hunchback producida (es decir, $[\text{Hb}] = [\text{Hb}]_{\text{max}} p_{\text{on}}$).

Llamamos ϵ_{off} y ϵ_{on} a las energías correspondientes a los estados off y on de cada sitio regulador, ϵ_{on} cuando Bcd no está ligado a ningún sitio. Si el gen hunchback está "activado", los sitios de unión para Bicoid pueden estar ocupados por hasta 6 moléculas de Bcd. Suponemos que la constante de disociación es la misma para todos los sitios (y que son independientes entre sí). Suponiendo que el peso correspondiente a cada sitio ocupado es:

$$1 + [\text{Bcd}]/K_{\text{d}}^{\text{on}}$$

Y suponiendo que los 6 sitios son independientes entre sí, se obtiene:

$$p_{\text{on}} = \frac{e^{-\beta\epsilon_{\text{on}}} (1 + [\text{Bcd}]/K_{\text{d}}^{\text{on}})^6}{e^{-\beta\epsilon_{\text{on}}} (1 + [\text{Bcd}]/K_{\text{d}}^{\text{on}})^6 + e^{-\beta\epsilon_{\text{off}}}}.$$

Y usando:

$$[\text{Bcd}] = [\text{Bcd}]_{\text{max}} e^{-x/\lambda},$$

Se relaciona p_{on} con x . Y con esta expresión es que se puede determinar con cuánta precisión es necesario conocer $[\text{Bcd}](x)$ para que la frontera bien aguda de Hb se pueda formar.

Una pequeña digresión sobre de dónde viene esto:

$$1 + [\text{Bcd}]/K_d^{\text{on}}$$

$$p_{\text{on}} = \frac{e^{-\beta \epsilon_{\text{on}}} (1 + [\text{Bcd}]/K_d^{\text{on}})^6}{e^{-\beta \epsilon_{\text{on}}} (1 + [\text{Bcd}]/K_d^{\text{on}})^6 + e^{-\beta \epsilon_{\text{off}}}}$$

Viene de pensar al sistema al que se le pueden unir muchas moléculas (en este caso, de Bcd) como un sistema que intercambia energía y partículas con el medio que lo rodea.

Entonces ahora uno se pregunta no solo cuál es la probabilidad de que el sistema esté en un micro estado de una dada energía sino también con un dado número de partículas “agregadas”.

En ese caso la función de partición es:

$$\mathcal{Z} = \sum_i e^{-\beta(E_s^{(i)} - N_s^{(i)} \mu)}$$

con μ el potencial químico que para un soluto (en este caso, Bcd) en una solución diluida es:

$$\mu_i = \mu_{i0} + k_B T \ln \frac{c_i}{c_{i0}}$$

y esto después se aplica al modelo:

con $c = [\text{Bcd}]$

Modelo MWC

$$\begin{array}{c}
 T_0 \\
 \downarrow \uparrow \\
 R_0 \xrightleftharpoons[3]{6\alpha c} R_1 \xrightleftharpoons[2\beta]{5\alpha c} R_2 \xrightleftharpoons[3\beta]{4\alpha c} R_3 \xrightleftharpoons[4\beta]{3\alpha c} R_4 \xrightleftharpoons[5\beta]{2\alpha c} R_5 \xrightleftharpoons[6\beta]{\alpha c} R_6
 \end{array}$$

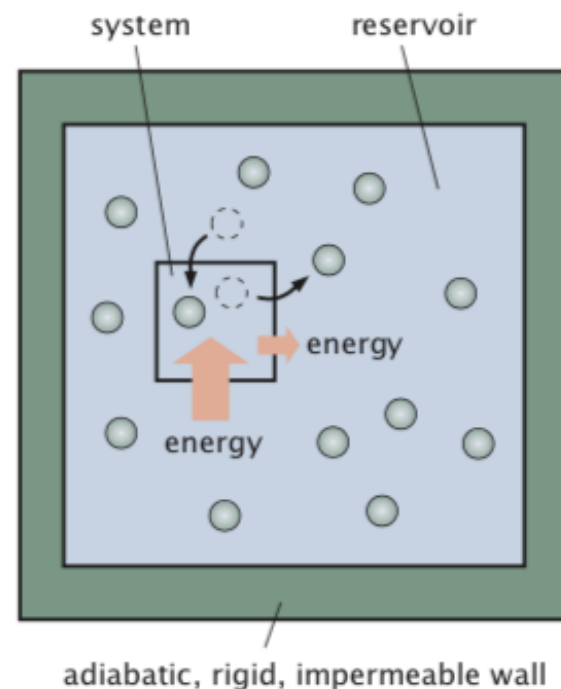


Figure 7.9: System in contact with a particle and energy reservoir. The system and reservoir are isolated from the rest of the world by adiabatic, rigid, and impermeable walls. However, the system and reservoir themselves can exchange both energy and particles.

Considerando dos núcleos “contiguos” a lo largo del eje, podemos suponer que la diferencia en la concentración de Bcd en ambos $\Delta[\text{Bcd}] = [\text{Bcd}](x) - [\text{Bcd}](x + \Delta x)$ es:

$$\frac{\Delta[\text{Bcd}]}{[\text{Bcd}]} = 1 - e^{-\Delta x/\lambda} \approx \frac{\Delta x}{\lambda} \approx 0.1$$

donde se tuvo en cuenta que $\lambda \sim 100\mu\text{m}$ y la separación entre núcleos contiguos es $\sim 10\mu\text{m}$

Esto significa que para que el salto agudo en Hb pueda ocurrir es necesario que se puedan distinguir concentraciones de Bcd que difieren en 10% en la zona donde ocurre el salto. En esa región es $[\text{Bcd}] = 5\text{nM}$, es decir ~ 500 moléculas de Bicoid en el núcleo.

Modelo más sencillo de ligado de Bcd al sitio regulador: $p_{\text{bound}} = \frac{\frac{[\text{Bcd}]}{K_d}}{1 + \frac{[\text{Bcd}]}{K_d}}$ de donde:

$$\frac{[\text{Bcd}]}{K_d} = \frac{p_{\text{bound}}}{1 - p_{\text{bound}}},$$

Y

$$\frac{\Delta[\text{Bcd}]}{[\text{Bcd}]} = \frac{\Delta p_{\text{bound}}}{\bar{p}_{\text{bound}}} + \frac{\Delta p_{\text{bound}}}{1 - \bar{p}_{\text{bound}}}.$$

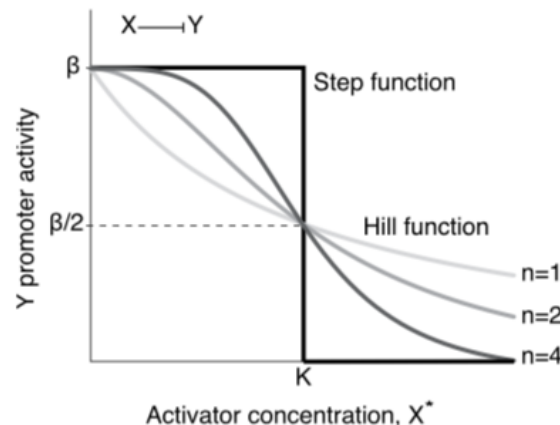
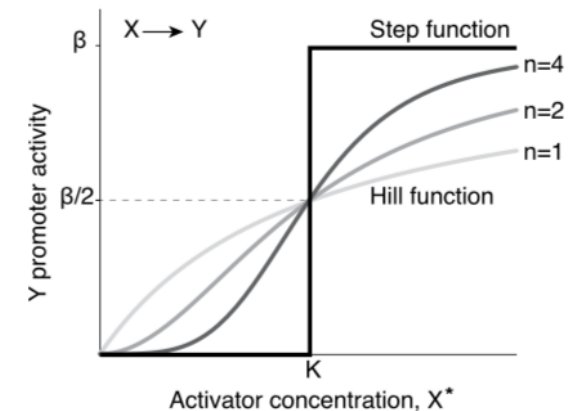
Por lo tanto, debe ser: $\Delta p_{\text{bound}}/p_{\text{bound}} = 0.05$ para que sea $\Delta[\text{Bcd}]/[\text{Bcd}] = 0.1$

Después está el tema del tiempo que lleva alcanzar esa precisión que ya vimos.

Hasta ahora vimos solo cálculos de equilibrio, que no tuvieron en cuenta la dinámica de la transcripción.

De algún modo lo que estuvimos viendo es la “fuerza” con la que un factor de transcripción afecta la expresión de uno de sus genes objetivo. Podemos pensar a esta “fuerza” como una función de entrada como se muestra en la figura (libro U. Alon) donde X es un factor de transcripción que regula la producción de la proteína Y . Cuando X regula Y , representado en la red por $X \rightarrow Y$, el número de moléculas de proteína Y producidas por unidad de tiempo es una función de la concentración de X en el estado activo, X^* .

Normalmente, la función de entrada $f(X^*)$ es una función monótona, creciente cuando X es un activador como en la figura y decreciente cuando X es un represor. Una función útil que describe de manera realista muchas funciones de entrada de genes es la función Hill. La función de Hill puede derivarse de considerar la unión de equilibrio del factor de transcripción a su sitio en el promotor y el índice, n , da cuenta de la cooperatividad de la unión.



Para un represor, $X \rightarrow Y$, la función de entrada de Hill es una curva decreciente

$$f(X^*) = \beta \frac{K^n}{K^n + X^{*n}}$$

Dado que un represor permite la transcripción fuerte de un gen solo cuando no está unido al promotor, esta función se puede derivar considerando la probabilidad de que el promotor no esté ligado por X^* .

En este nivel de descripción se puede agregar a las flechas de las redes los números que describen estas funciones de Hill, β , K y n .

Estos parámetros pueden variar con la evolución. Por ejemplo, K puede cambiarse por mutaciones que alteran la secuencia de ADN del sitio de unión de X en el promotor. Incluso un cambio de una sola letra de ADN en el sitio de unión puede fortalecer o debilitar los enlaces químicos entre X y el ADN y cambio de K . El parámetro K también puede variar si cambia la posición del sitio de unión, así como por algunos cambios en la secuencia fuera del sitio de unión. De manera similar, la actividad máxima β puede variar por mutaciones en el sitio de unión de la RNAP.

Se hacen experimentos de evolución en el laboratorio que muestran que cuando las bacterias son colocadas en un nuevo entorno pueden ajustar con precisión estos parámetros a lo largo de las generaciones para alcanzar niveles óptimos de expresión. Por lo tanto, estos números están bajo presión de selección y pueden cambiar de manera heredable a lo largo de muchas generaciones si los entornos cambian.

Las funciones de Hill son útiles para modelos detallados. Sin embargo, a veces es útil usar funciones aún más simples que capturan el comportamiento esencial de las funciones de entrada. Su esencia es la existencia de una transición entre valores bajos y altos, con un umbral característico K . Una aproximación sencilla (la aproximación lógica) es suponer que el gen solo puede estar desactivado, $f(X^*) = 0$, o tener una activación máxima, $f(X^*) = \beta$ y que el umbral de transición ocurre para $X^* = K$. Para activadores y represores será:

$$f(X^*) = \beta \theta(X^* > K)$$

$$f(X^*) = \beta \theta(X^* < K)$$

Respectivamente, con θ la función escalón.

Esta descripción en términos de funciones lógicas permite combinar regulaciones como si se tratara de compuertas lógicas.

Por ejemplo, supongamos un gen regulado por dos activadores. Muchos genes requieren que ambas proteínas activadoras se unan al promotor para mostrar una alta expresión. Esto es similar a una compuerta AND:

$$f(X^*, Y^*) = \beta \theta(X^* > K_x) \theta(Y^* > K_y) \sim X^* \text{ AND } Y^*$$

Para otros genes, la unión de cualquiera de los activadores es suficiente. Esto se asemeja a una compuerta OR:

$$f(X^*, Y^*) = \beta \theta(X^* > K_x \text{ OR } Y^* > K_y) \sim X^* \text{ OR } Y^*$$

Sin embargo, no todos los genes tienen funciones de entrada de tipo booleano. Por ejemplo, algunos genes muestran una función de entrada que es la suma de las individuales (Kalir y Alon, 2004):

$$f(X^*, Y^*) = \beta_x X^* + \beta_y Y^*$$

Y también se pueden encontrar otras funciones con mesetas y valles como en la figura que corresponde a operón lac de E. coli.

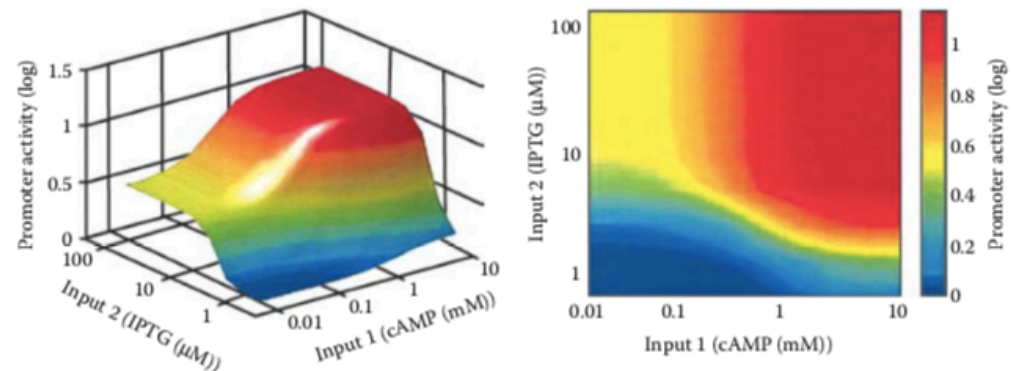


FIGURE 1.8 Adapted from (Setty et al., 2003).

Vemos entonces que hay muchos niveles de descripción.Cuál es el que interesa deberá relacionarse con las preguntas que busquemos contestar y con la posibilidad de relacionar con experimentos.

Y encima todavía no nos metimos con la dinámica (lo haremos brevemente en otra presentación). Y menos con la parte espacial.

Antes de cerrar la presentación los dejo con una tabla que muestra la fuerte separación de escalas de tiempo que caracteriza a las redes transcripcionales lo que explica por qué tiene sentido trabajar con relaciones de equilibrio muchas veces: las señales de entrada generalmente cambian las actividades de los factores de transcripción en una escala de tiempo inferior a un segundo. La unión del factor de transcripción activo a sus sitios de ADN a menudo alcanza el equilibrio en segundos. La transcripción y traducción del gen “objetivo” lleva unos minutos y la acumulación del producto proteico puede llevar horas. Por tanto, los diferentes pasos entre la señal de entrada y la acumulación de la proteína de salida tienen escalas de tiempo muy diferentes.

TABLE 1.2 Typical Timescales for the Reactions in the Transcription Network of *E. coli*

Binding of a small molecule signal to a transcription factor, causing a change in its activity	~1 msec
Binding of an active transcription factor to its DNA site	~1 sec
Transcription and translation of the gene	~minutes
Timescale for 50% change in concentration of the translated protein (stable proteins)	~1 hour (1 cell generation)

Por lo tanto, se puede considerar que los niveles de actividad de los factores de transcripción están en un estado cuasi-estacionario de las ecuaciones dinámicas que describen los cambios en los niveles de proteínas en la escala lenta. Además de las redes de transcripción, la célula contiene varias otras redes de interacciones, como las redes de transducción de señales compuestas por proteínas que interactúan, que operan mucho más rápido que las redes de transcripción y, por lo tanto, se pueden considerar en un estado cuasi-estacionario también.

Los aspectos dinámicos, sin embargo, no pueden ser dejados de lado así nomás, ya que el “mensaje” que tienen que decodificar las células u organismos a veces están “codificados” en la concentración de las sustancias mensajeras y en otras circunstancias en su dinámica temporal.

Una propiedad adicional notable de las redes de transcripción es la modularidad de sus componentes. Así se puede tomar el ADN de un gen de un organismo y expresarlo en un organismo diferente, lo que se usa para fines experimentales, e.g., para ver los niveles de proteína in vivo usando agregando el gen de la GFP a la secuencia de interés.

Los componentes modulares hacen que las redes de transcripción pueden evolucionar con bastante rapidez (aquí "rápidamente" se refiere a la escala de muchas generaciones): las regulaciones que caracterizan a las redes de transcripción evolucionan en una escala de tiempo mucho más rápida que las regiones codificantes de los genes. Por ejemplo, los ratones y los seres humanos tienen genes muy similares, pero la regulación de la transcripción de estos genes, que gobierna cuándo y qué cantidad de cada proteína se produce, es diferente. En otras palabras, muchas de las diferencias entre las especies animales radican en las diferencias en la regulación de las redes de transcripción, más que en las diferencias en sus genes.

Esta modularidad, por otro lado, llama a analizar qué aspectos tienen en común las redes que regulan la expresión en distintos organismos.

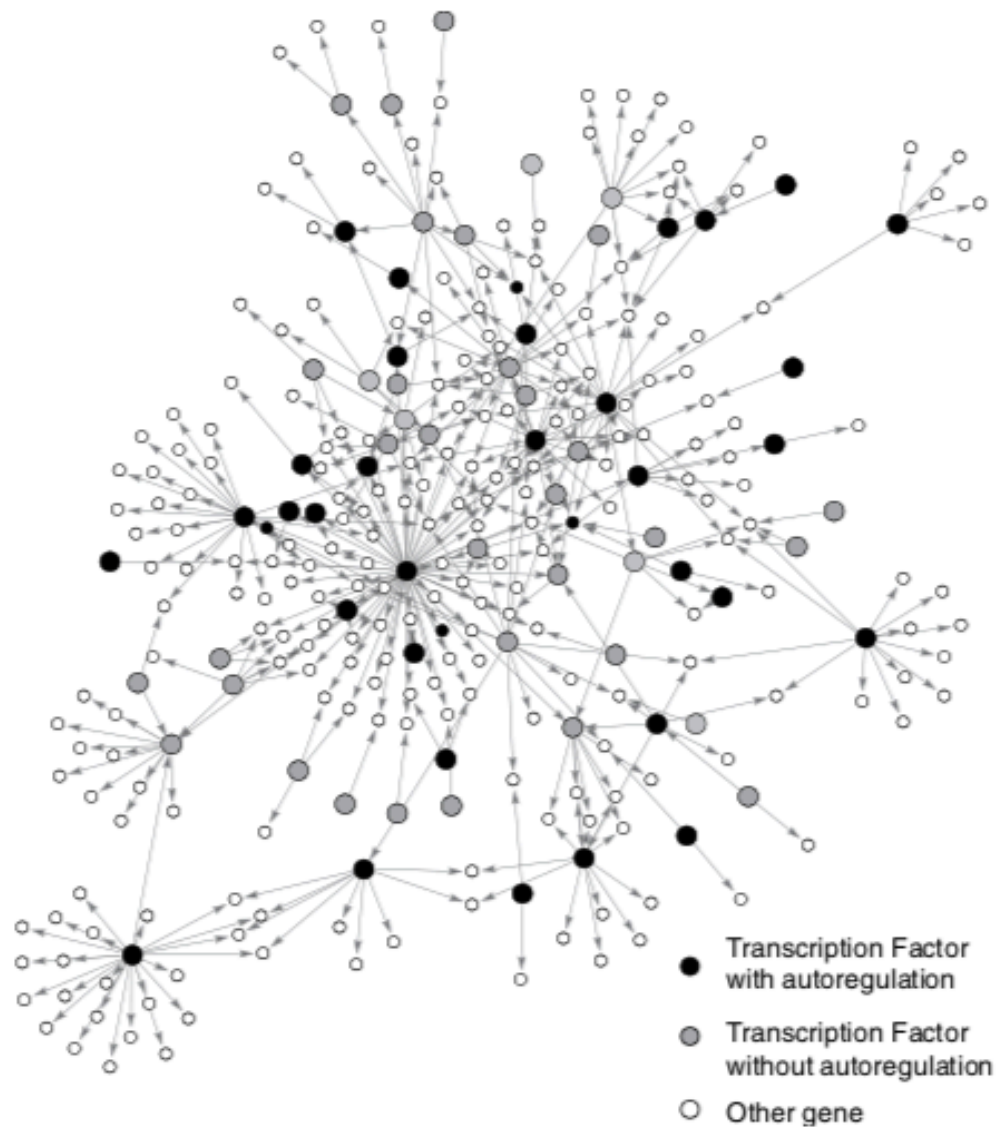
Esta visión es la que subyace, de algún modo, a los análisis que se hacen usando ideas de teoría de redes en los que se busca si hay “estructuras” (motivos) que aparecen con mayor frecuencia que otros.

Un “motivo” es una estructura de conexiones que involucra a pocos nodos de la red pero que se repite casi como si fuera un “ladrillo” de una construcción más complicada.

Para encontrar “motivos” prevalentes es necesario comparar la red verdadera con una red aleatoria que tenga las mismas características que la red real (por ejemplo, el mismo número de nodos y flechas que la red real), pero donde las conexiones entre nodos y flechas se realizan al azar. Los patrones que ocurren en la red real con mucha más frecuencia que en las redes aleatorias se denominan motivos de red.

Un motivo recurrente es la auto-regulación negativa.

Esto se puede ver, por ejemplo, en la red de *E. coli*. Esta red tiene $N=424$ nodos y $A=519$ flechas. Entre estas últimas hay 40 “auto-flechas” (que salen y apuntan a los factores de transcripción que se auto-regulan). Una red al azar con este número de nodos y flechas debería tener 1 ± 1 auto-flechas.



¿Por qué la autorregulación negativa es un motivo de red? ¿Tiene alguna función útil?

Una de sus funciones es permitir que la regulación en la concentración de las proteínas producidas pueda ocurrir más rápidamente que si no existiera dicha la regulación negativa.

Para eso, tenemos que ver algo sobre la dinámica de la transcripción.