

Una introducción a la Biología Molecular

Para Computadores Científicos
interesados en Bioinformática

Fernán Agüero,

Instituto de Investigaciones Biotecnológicas, UNSAM

<fernan@iib.unsam.edu.ar>

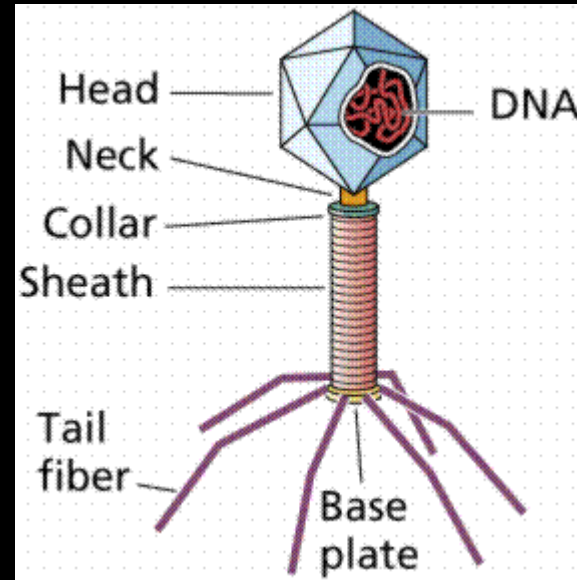
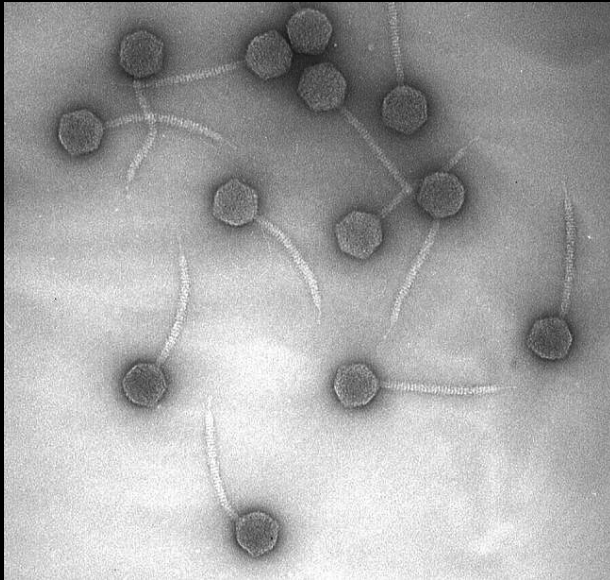
Referencias

- Molecular Biology for Computer Scientists. Lawrence Hunter, Chapter 1 in Artificial Intelligence and Molecular Biology (L. Hunter, ed.), 1993.
- <http://www.aaai.org/Library/Books/Hunter/01-Hunter.pdf>
- A quick introduction to elements of biology - cells, molecules, genes, functional genomics, microarrays. Alvis Brazma, Helen Parkinson, Thomas Schlitt, Mohammadreza Shojatalab, 2001.
- http://www.ebi.ac.uk/microarray/biology_intro.html
- Online Book of Biology. MJ Farabee, 1992-2002.
- <http://www.emc.maricopa.edu/faculty/farabee/BIOBK/BioBookTOC.html>
- Molecular Biology of the Cell, Alberts *et al* 4th Ed, 2002, Garland Publishing.
- <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?call=bv.View..ShowTOC&rid=mboc4.TOC&depth=2>

Seres vivos: algunas definiciones

- Características centrales que nos permiten decir que algo “es un ser u organismo vivo”:
 - Capacidad de reproducirse a sí mismo
 - Toma materiales del entorno y utiliza energía que captura del entorno o de estos materiales para transformarlos en componentes propios.

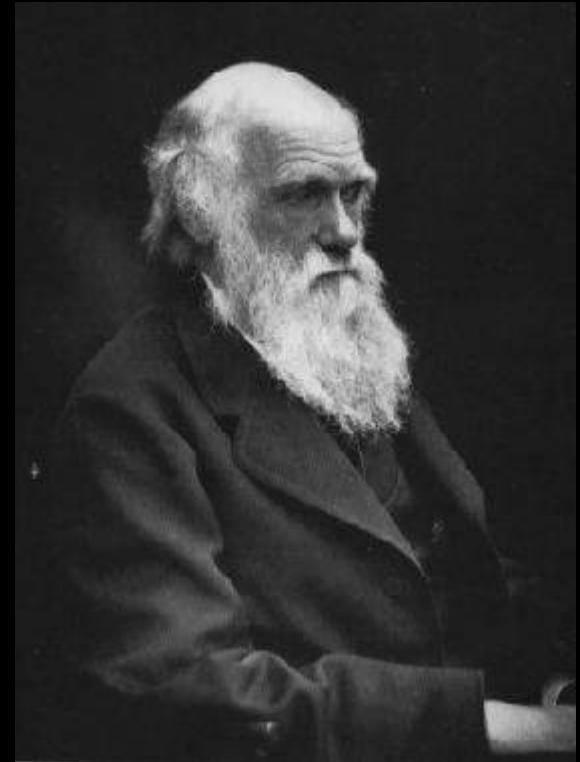
En los límites: virus



Bacteriofago Lambda – virus que usan bacterias como hospedadores. Izq: microscopía electrónica. Der: esquema

Todos los seres vivos están relacionados entre sí

- Cualquier par de organismos tienen un ancestro en común
- Las diferencias entre organismos y la adquisición de la complejidad que observamos pueden ser explicadas mediante la *evolución*
 - Herencia
 - Variación
 - Selección
- Estos tres factores definen un proceso evolutivo



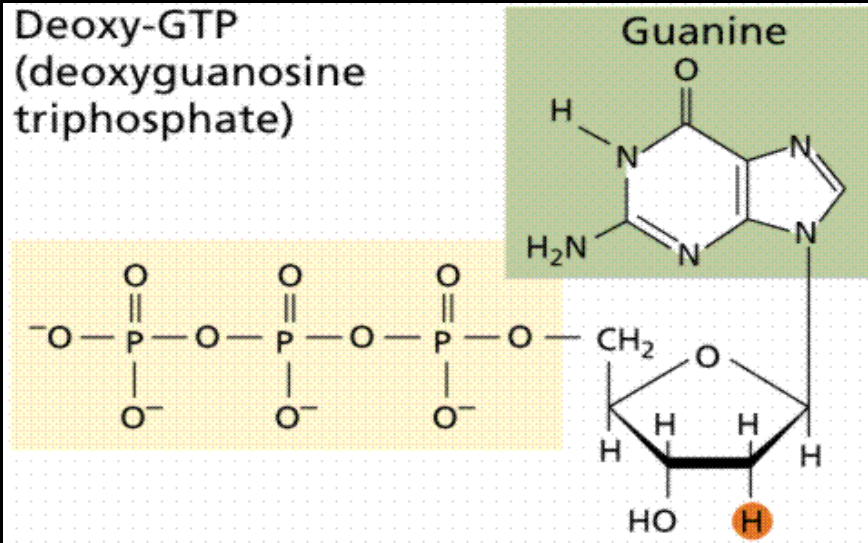
Charles Darwin

Herencia

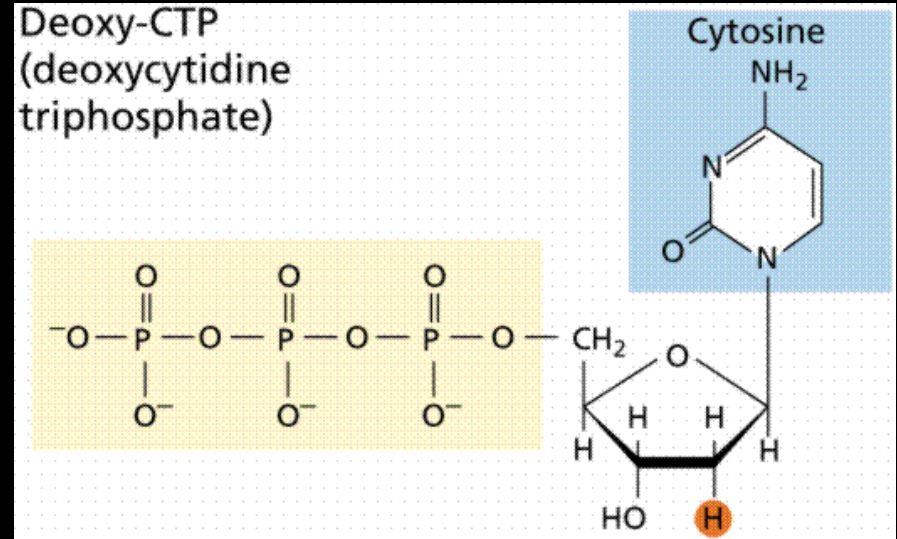
- Toda la información que se hereda se encuentra almacenada en forma química en una molécula: el DNA
 - Es un polímero lineal
 - Construído a partir de cuatro monómeros: adenina, citosina, guanina y timidina
 - mas conocidos como A, T, C y G :)

Componentes del DNA

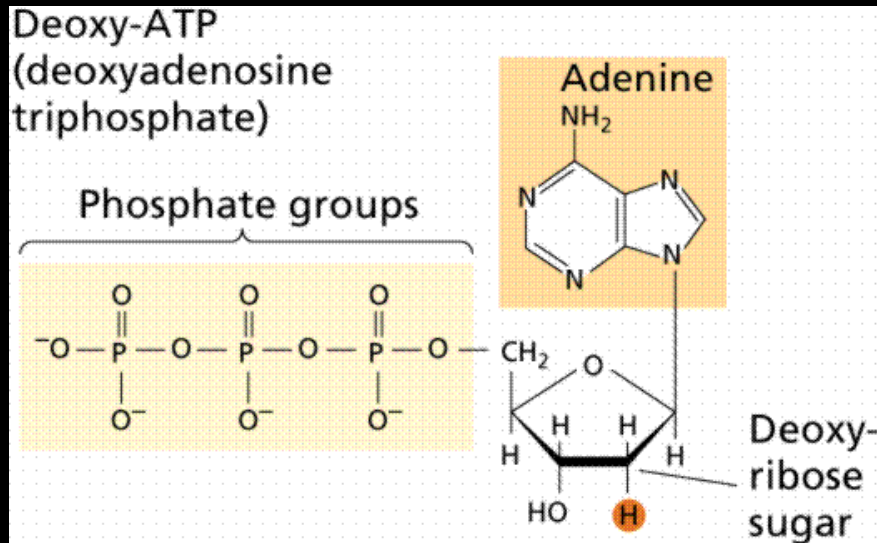
Deoxy-GTP
(deoxyguanosine
triphosphate)



Deoxy-CTP
(deoxycytidine
triphosphate)

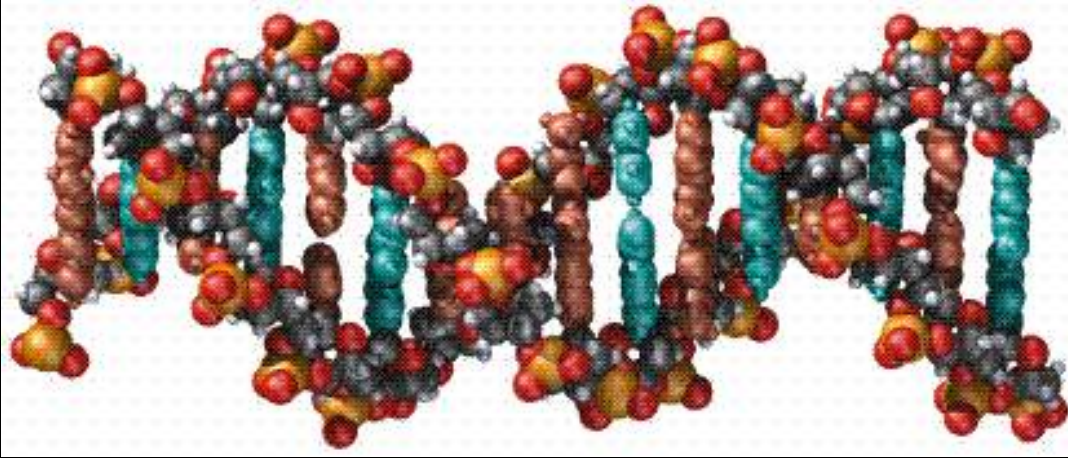


Deoxy-ATP
(deoxyadenosine
triphosphate)

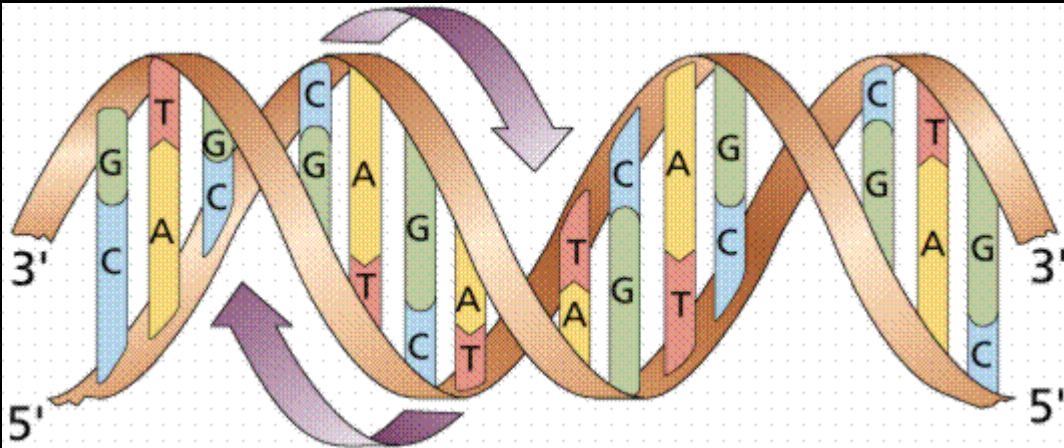


- Base nitrogenada
- Azucar (deoxyribose)
- Fosfato
- Observar:
 - OH 3' libre
 - PO₄ en posición 5'

Estructura del DNA

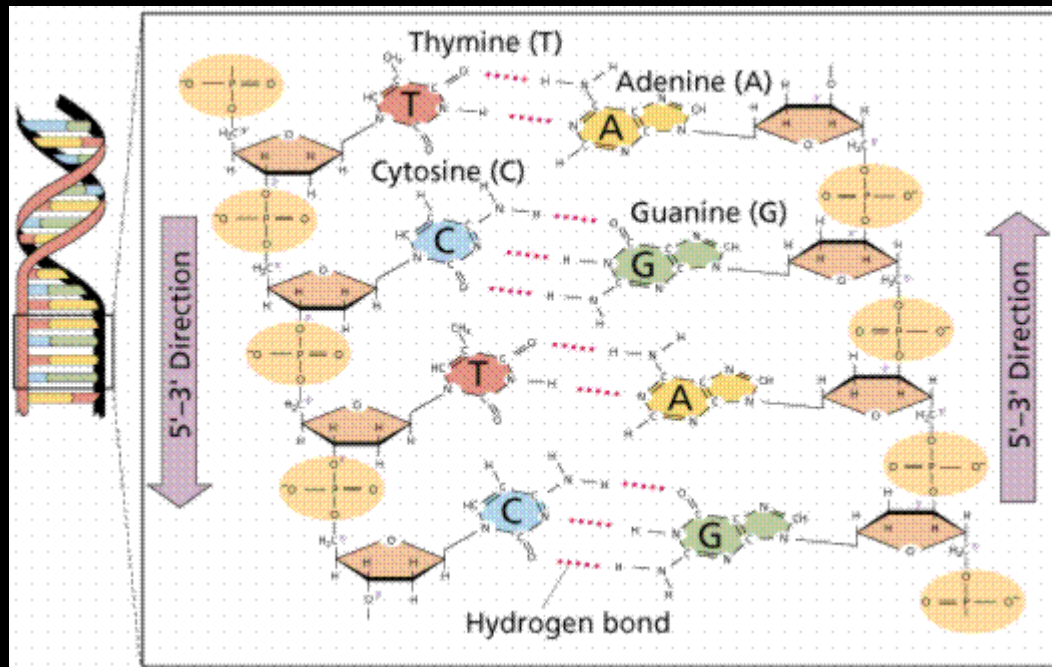


Modelo atómico



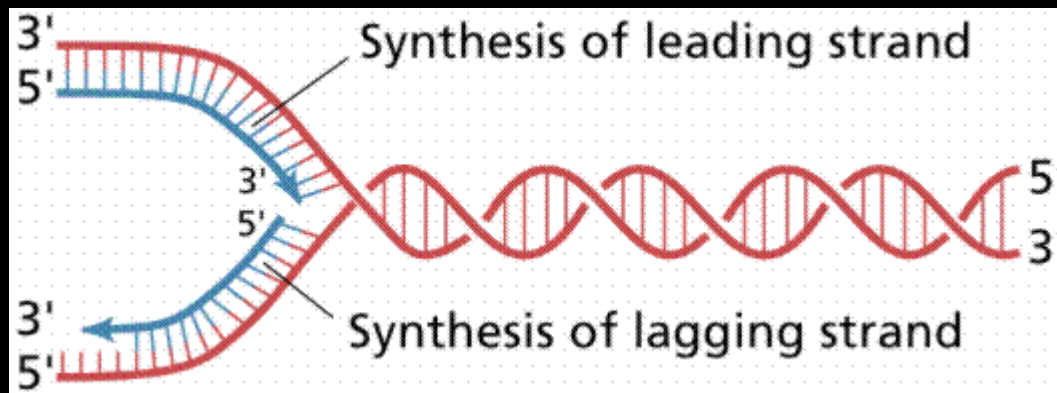
Esquema

Estructura del DNA: detalle



Replicación del DNA

- La replicación del DNA es semi-conservativa
 - La doble hélice se abre y cada hebra sirve como molde para la síntesis de dos nuevas hebras



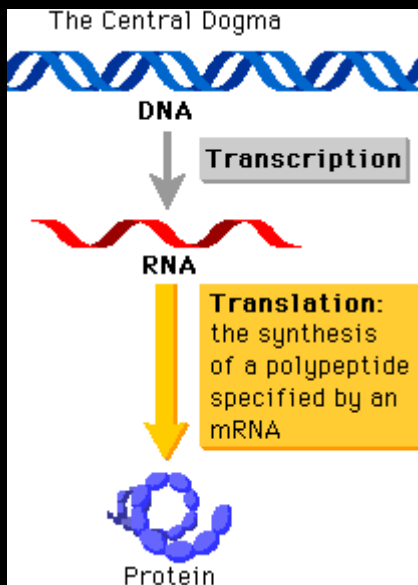
La síntesis del DNA durante la replicación

- Ocurre en una única dirección: 5' a 3'
- Es diferente en cada hebra:
 - Leading strand (hebra líder): síntesis continua 5' a 3'
 - Lagging strand (hebra retrasada): síntesis discontinua 5' a 3'
- Es enzimática
 - Complejos macromoleculares multifuncionales
 - Algunas funciones son:
 - Polimerización
 - Corrección (eliminación de bases mal apareadas)
 - Desplegado y apertura de la doble hélice
 - Ligado de los fragmentos de síntesis en la hebra retrasada (lagging strand)

Codificación de información

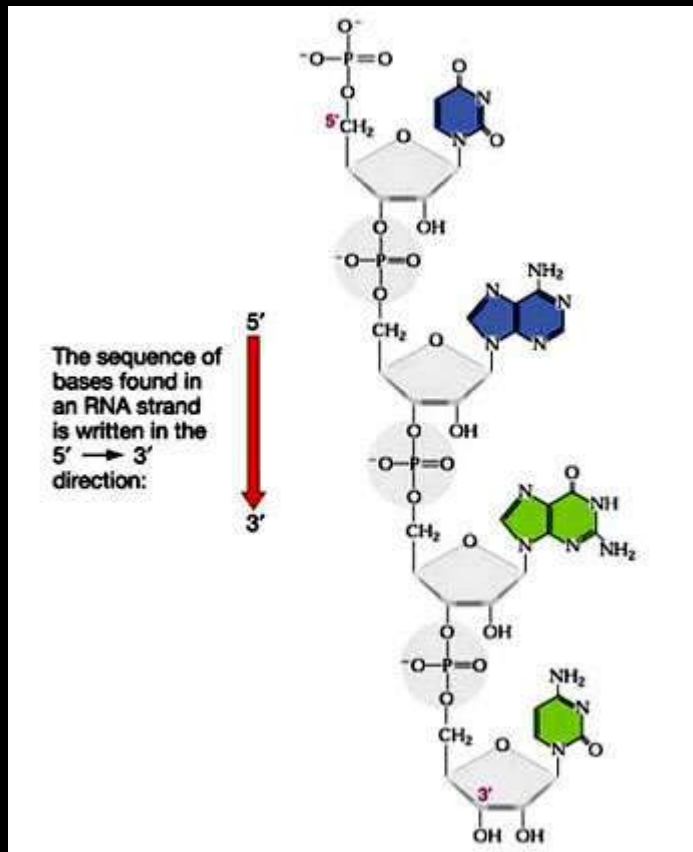
- Almacenamiento de información: DNA
- Alfabeto de 4 caracteres
- Cómo está codificada esta información (es decir, como leemos lo que sea que está allí para obtener algún mensaje con sentido)
 - Respuesta: necesitamos conocer cómo es el flujo de la información en biología

El dogma central de la biología molecular



El dogma central de la biología molecular describe el proceso en dos pasos (transcripción y traducción) por el cual la información genética lleva a la producción de proteínas: $\text{DNA} \rightarrow \text{RNA} \rightarrow \text{proteína}$.

RNA



- El RNA es también un ácido nucleico
 - También se polimeriza a partir de nucleótidos, en general copiando a partir de un molde.
- Diferencias?
 - Tiene dos OH libres (2' y 3') (es decir está formado por ribosa, en lugar de desoxyribosa)
 - En el RNA se utiliza uracilo como base nitrogenada en lugar de timina). Es decir el alfabeto sería: A, C, G, U

Transcripción

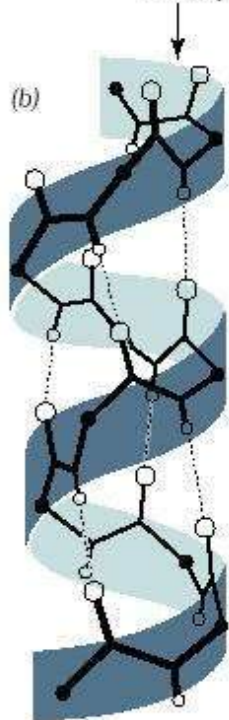
- Conceptualmente similar a la replicación del DNA
 - Se utiliza DNA como molde
 - Pero se polimeriza una hebra de RNA
- Pero
 - No se transcribe todo el DNA
 - Existen señales en la secuencia de DNA que delimitan de alguna manera la región a transcribir
 - La maquinaria celular encargada de realizar la transcripción es la que detecta y responde a estas señales
- En el proceso de transcripción se mantiene la información contenida en el DNA. Sólo cambia la forma química en la que está almacenada.

Traducción

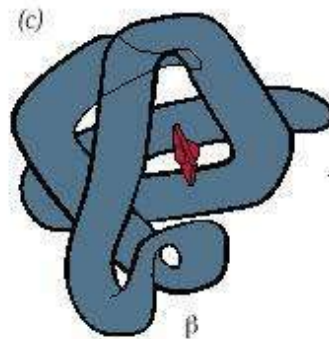
- Proteínas
 - Polímeros de aminoácidos
 - Alfabeto: 20 caracteres (20 aminoácidos)
 - El polímero se pliega en el espacio formando una estructura tridimensional par
- Funciones realizadas por proteínas
 - Catálisis de reacciones químicas (enzimas)
 - Soporte estructural (citoesqueleto)
 - Transporte de moléculas
 - Otras
-

Proteínas: la función está determinada por la estructura

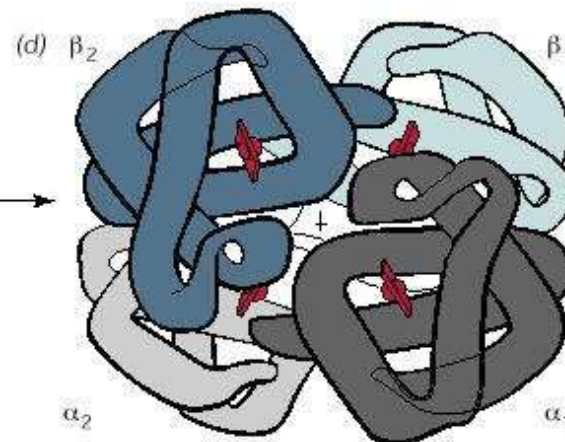
(a) \pm Lys \pm Ala \pm His \pm Gly \pm Lys \pm Lys \pm Val \pm Leu \pm Gly - Ala \pm
Primary structure (amino acid sequence in a polypeptide chain)



Secondary structure (helix)



Tertiary structure:
one complete protein chain
(β chain of hemoglobin)



Quaternary structure:
the four separate chains
of hemoglobin assembled
into an oligomeric protein

Figure 6-1. Levels of protein structure.

(a) Primary structure, (b) secondary structure, (c) tertiary structure, and (d) quaternary structure. [Figure copyrighted © by Irving Geis.]

Traducción: el código genético

- Cómo traducimos mensajes escritos usando un alfabeto de 4 caracteres (DNA/RNA) en uno de 20?
 - La maquinaria celular lo hace utilizando tripletes de bases (codones)
 - La tabla de correspondencias entre todas las posibles secuencias de DNA de tres letras y los aminoácidos que codifican es lo que se llama el código genético
 - Hay 64 combinaciones posibles utilizando 4 letras tomadas de a 3
 - Pero sólo hay 20 aminoácidos:
 - El código genético es redundante (degenerado). Más de un triplete codifica para un mismo aminoácido

El código genético

		Second letter				
		U	C	A	G	
U	UUU } Phe	UCU } Ser	UAU } Tyr	UGU } Cys	U	
	UUC } Leu		UAC } Stop			UGC } Stop
	UUA } Leu		UAA } Stop			UGA } Stop
	UUG } Leu		UAG } Stop			UGG } Trp
C	CUU } Leu	CCU } Pro	CAU } His	CGU } Arg	U	
	CUC } Leu		CAC } Gln			CGC } Arg
	CUA } Leu		CAA } Gln			CGA } Arg
	CUG } Leu		CAG } Gln			CGG } Arg
A	AUU } Ile	ACU } Thr	AAU } Asn	AGU } Ser	U	
	AUC } Ile		AAC } Lys			AGC } Ser
	AUA } Met		AAA } Lys			AGA } Arg
	AUG } Met		AAG } Lys			AGG } Arg
G	GUU } Val	GCU } Ala	GAU } Asp	GGU } Gly	U	
	GUC } Val		GAC } Glu			GGC } Gly
	GUA } Val		GAA } Glu			GGA } Gly
	GUG } Val		GAG } Glu			GGG } Gly

•3 aminoácidos son codificados por 6 tripletes:

•Serina (Ser, S), Arginina (Arg, R), Leucina (Leu, L)

•5 por 4 tripletes:

•Valina (Val, V), Alanina (Ala, A), Glicina, (Gly, G), Prolina (Pro, P), Treonina (Thr, T)

•1 por 3 tripletes:

•Isoleucina (Ile, I)

•9 por 2 tripletes:

•Fenilalanina (Phe, F), Tirosina (Tyr, Y), Cisteina (Cys, C), Histidina (His, H), Glutamina (Gln, Q), Asparragina (Asn, N), Lisina (Lys, K), Aspartico (Asp, D), Glutamico (Glu, E)

•2 por 1 triplete:

•Metionina (Met, M), Triptofano (Trp, W)

•3 tripletes no codifican aminoácido alguno (fin de la traducción)

Lectura de tripletes

- Dada una secuencia de DNA proveniente de un gen, hay 3 sitios posibles a partir de los cuales iniciar la lectura en términos de aminoácidos (traducción)
- ...AATGCGATAAG...
 - Puede ser leída como AAT-GCG-ATA-AG...
 - Como ...A-ATG-CGA-TAA-G...
 - O como ...AA-TGC-GAT-AAG...
- Es decir hay 3 posibles marcos de lectura

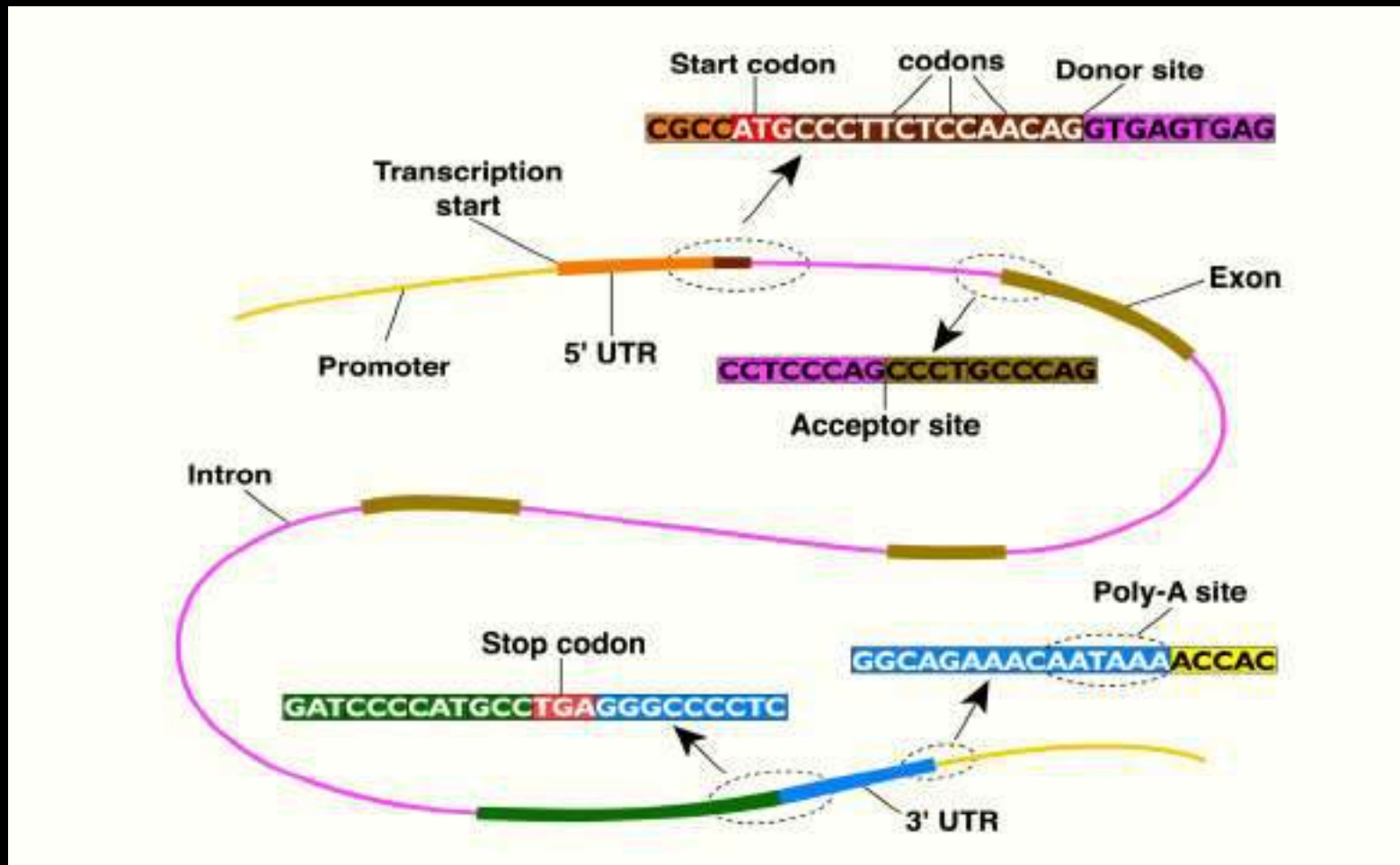
Estructura de un gen procariótico

GENE SEQUENCE	PATTERN
1 GAATTCGATAAAATCTCTGGGTTTATTGTCGCACTTTATGGTTT	CTGNNNNNNNNNNNCAG
	TTGACA
41 CC AAAATCGCCTTTTTCCTGTATATATACATCACAGCATAACTG	CTGNNNNNNNNNNNCAG
CGAA -35 -10 TATACT >	TATAAT, > mRNA start
81 TATAATCACCCAGGGGGCGGAAATGAAAGCGTTAAACGGCCA	CTGNNNNNNNNNNNCAG
+10 GGGGG Ribosomal binding site	GGAGG
121 GGCACAAGAGGTGTGTGATCTTCATCCGTGATCACATCAG	ATG
161 CCAGACAGGTATCCCGCCGACGCGTGCAGAAATCGGCGAG	
201 CGTTTGGGGTTCGGTCCCCAAACGGGCTGAAGAACATC	
241 TGAAGGCGCTGSCACGCAAGGCGTTATTGAAATGTGTTTC	
281 CGGCGCATCAGCGCGGTTTCGTCTGTTTCAGGAAGAGGAA	
321 GAAGGGTTGCGGCTGGTAGGTCGTGTGCGCTGCCCGTGRAC	
361 CACTTCTGGCGCAACAGCATATTTGAAGGTCATATCAGGT	OPEN READING FRAME
401 CGATCCCTTCCTTATTCAGCGGAAATGCTGATTTCCCTGCTG	
441 CGCGTCAGCGGGATGTGATGAAAATATTCGGCATTATGG	
481 ATGGTGTACTTGTCTGCGAGTGCATAAACTCAGGATGTACG	
521 TAACGGTCAGGTCTGTTGTCGCACATATTGATGACGAAGTT	
561 ACCGTTAAGCGCTGAAAAAACAGGGCAATTAAGTCGAAAC	
601 TGTTGCCAGAAAATAGCGAGTTTAAACCAATTGTCGTTGA	
641 CCTTCGTCAGCAGAGCTTCACCAATTAAGGGCTGGGGTT	TAA
681 GGGGTTATTGCAACGGCGACTGGCTGTACATATCTCTG	
721 AGACCGCGATGCGCGCTGGCGTTCGGGTTTGTTTTATC	
761 TCTCTTCATCAGGCTTGTCTGCATGGCATTCTCCTTCACA	
801 TCTGATAAAGCACTCTGGCATCTCGCCCTTACCCATGATTT	
841 TCTCCAATATCACCGTTCCGTTGCTGGGACTGTCGATAC	
881 GCGGGTAAATGGTCACTTGTATGACCCGTTTATTGTTGGC	
921 GCGGTGGCGGTTGGCGCAACGGGGACCAAGCT	

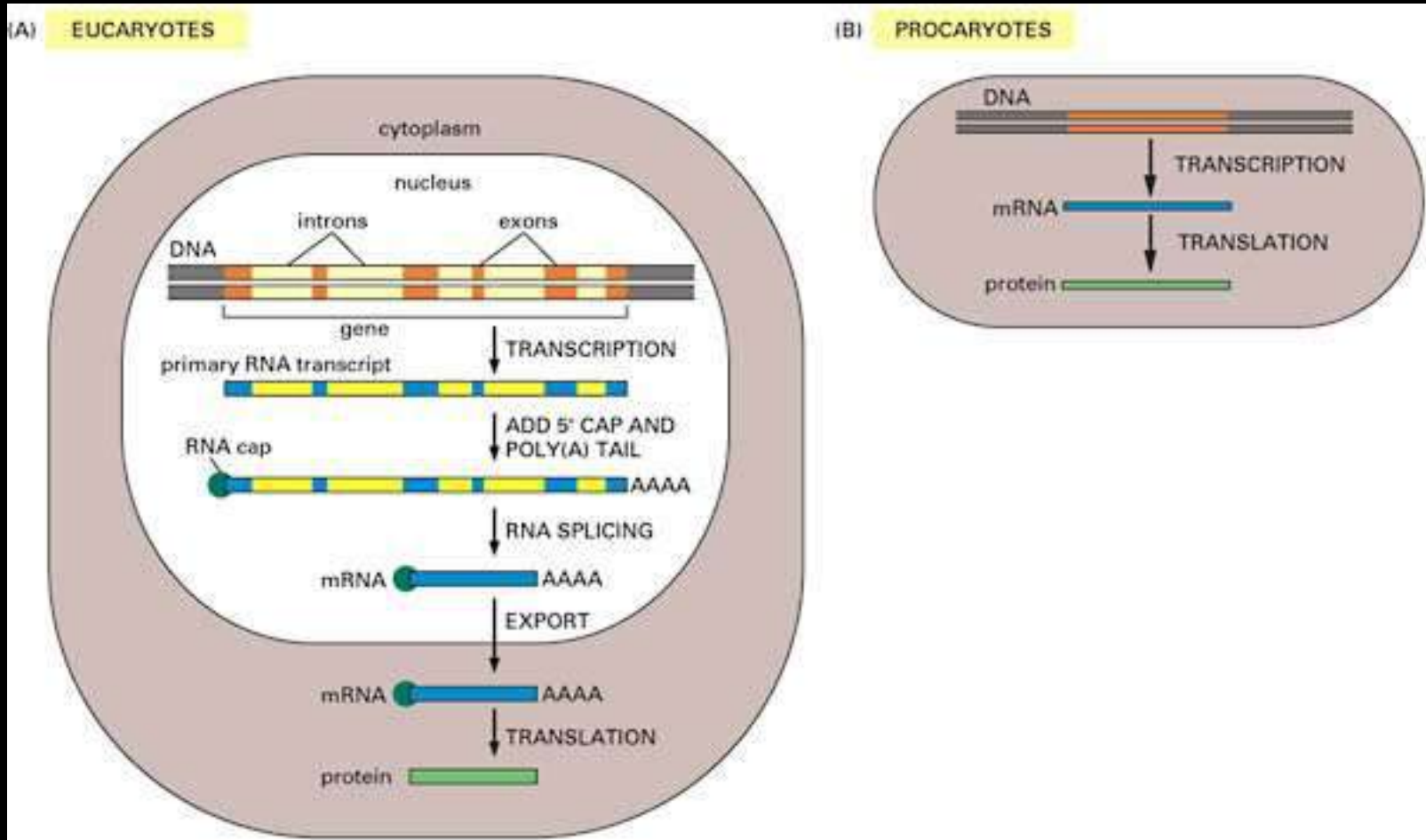
Shown are matches to approximate consensus binding sites for LexA repressor (CTGNNNNNNNNNNNCAG), the -10 and -35 promoter regions relative to the start of the mRNA (TTGACA and TATAAT), the ribosomal binding site on the mRNA (GGAGG), and the open reading frame (ATG...TAA). Only the second two of the predicted LexA binding sites actually bind the repressor.

- Secuencias regulatorias en la región 5' (no transcripta)
 - Median la unión de factores necesarios para la transcripción
 - Definen el sitio de inicio de la transcripción
- Secuencias regulatorias en la región 5' (transcripta, pero no traducida)
 - Median la unión del RNA al ribosoma para su traducción
- Puede haber otras secuencias regulatorias en los extremos 5' y 3' del mRNA que pueden afectar la traducción o su estabilidad

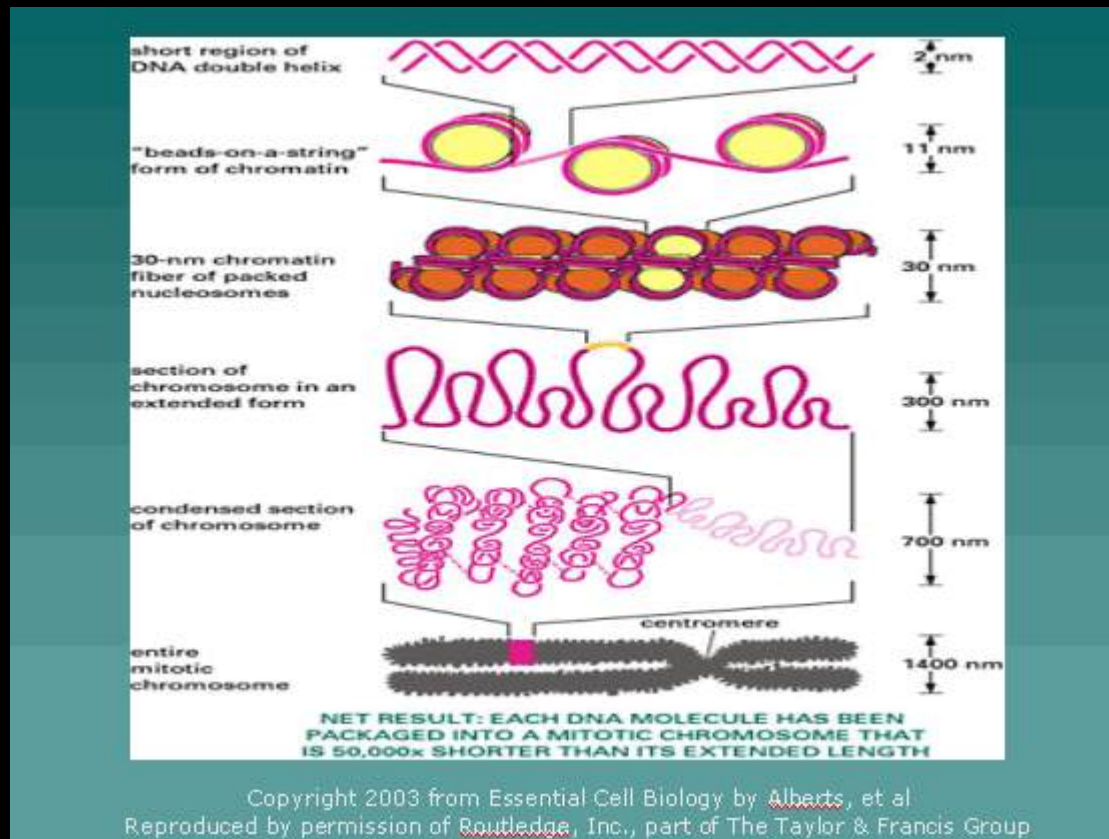
Estructura de un gen eucariótico



Estructura comparada del gen



Organización del DNA en el núcleo



El DNA en el laboratorio

- Clonado
- Hibridación
- Reacciones enzimáticas
- Aplicaciones:
 - Secuenciación
 - Southern blots
 - Microarrays

Qué hace la biología molecular

- También llamada tecnología del DNA recombinante o ingeniería genética
 - Aislar fragmentos de DNA de interés
 - unirlos en nuevas combinaciones
 - e introducir el nuevo DNA (recombinado) en un organismo vivo

Algunos conceptos centrales y definiciones

- La producción de DNA recombinante se realiza mediante la incorporación de un fragmento de interés en una molécula de DNA pequeña, que es capaz de ser replicada rápidamente
- Un organismo que contiene un fragmento de DNA introducido en forma artificial se dice que es transgénico
- El organismo a partir del cual se obtuvo el fragmento de DNA es el dador (donor)
- La molécula de DNA en la cual se inserta el fragmento de interés se llama vector
- Para obtener un fragmento de DNA a partir de un organismo dador se utilizan enzimas de restricción (funcionan como tijeras moleculares, cortando el DNA en sitios específicos)
- El DNA obtenido de esta forma, puede ligarse al vector utilizando una enzima particular (llamada DNA ligasa)

Mas definiciones

- Un vector conteniendo DNA de otro origen insertado en él se dice que representa una molécula de DNA recombinante (o quimérica)
- Los vectores más comunes son plásmidos que pueden replicar dentro de una bacteria
- El vector (DNA) y la bacteria (la célula hospedadora) se mezclan. El proceso por el cual la bacteria toma el DNA y lo internaliza se llama transformación.
- Las bacterias se dividen en forma clonal (las dos células resultantes de una división son idénticas)
- Si la bacteria contiene una molécula de DNA recombinante (por ejemplo un plásmido conteniendo un gen de interés), el aislamiento de una (una) célula (clon) nos permite decir que hemos clonado un gen

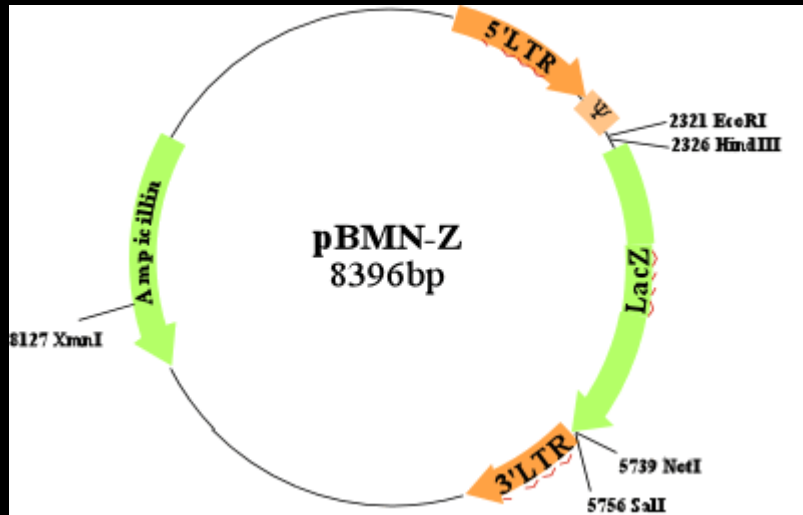
Qué es clonar?

- La biología molecular depende de técnicas *in vitro* e *in vivo* para poder manipular el DNA
 - Replicación en bacterias
 - Las bacterias poseen un único cromosoma circular
 - Algunas bacterias poseen además moléculas de DNA circulares más pequeñas: plásmidos
 - Los plásmidos existen dentro de la bacteria en varias copias por célula
 - Es posible manipular los plásmidos para insertar dentro de ellos algún fragmento de DNA de otro origen
 - Cuando las bacterias crecen en un medio semi-sólido (no en solución) es posible obtener colonias aisladas. Una colonia contiene millones de bacterias, todas provenientes de una única célula original.
 - Si esa colonia de bacterias alberga (en un plásmido, por ej) un gen de interés, decimos que hemos clonado ese gen.

Plásmidos

Mapa del plásmido pBMN-Z:

- Genes en verde
 - Resistencia a ampicilina
 - Capacidad de metabolizar lactosa

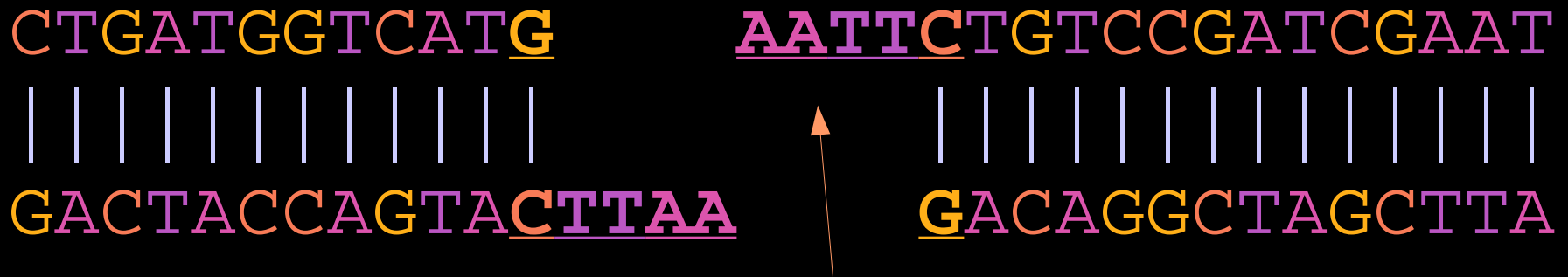
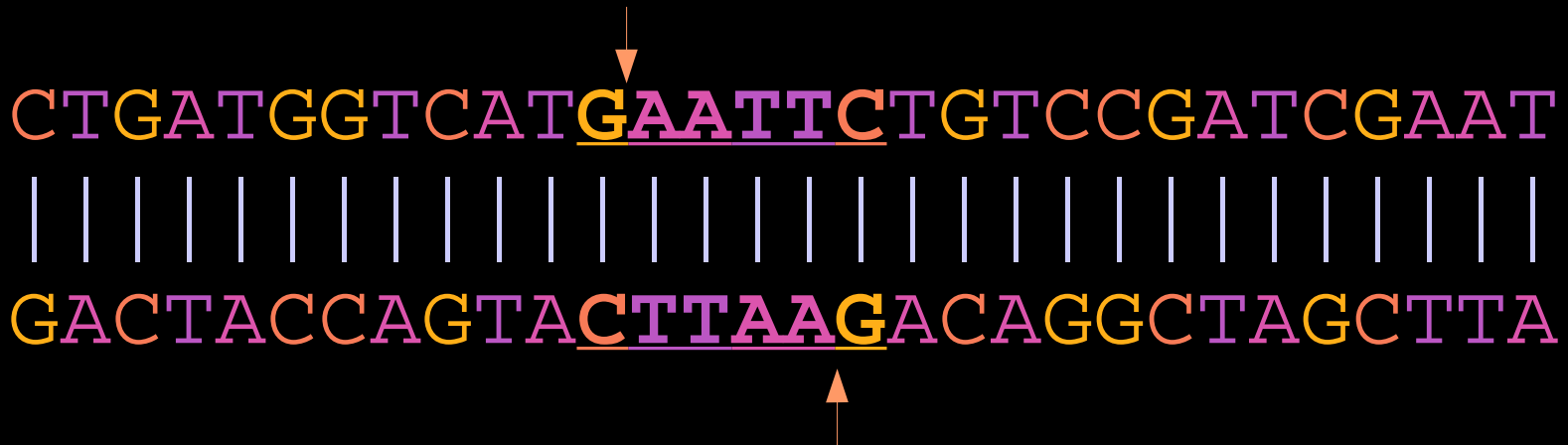


Manipulación *in vitro* del DNA

- Es posible realizar cortes en sitios específicos en una molécula de DNA
 - Enzimas de restricción
 - Son endonucleasas (cortan enlaces internos)
 - Reconocen secuencias específicas de distinta longitud
 - Cortan el DNA si hubo reconocimiento

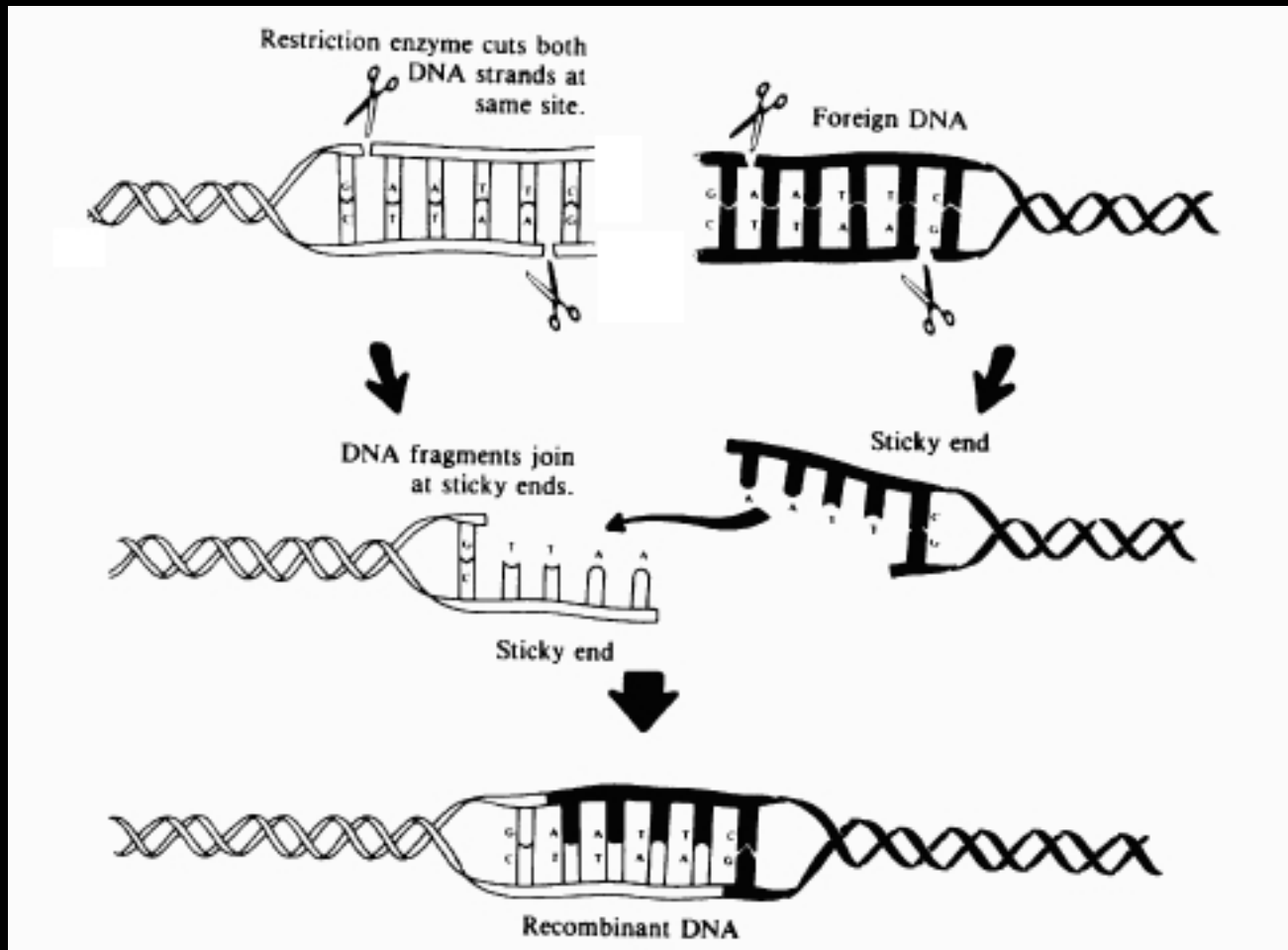
Enzimas de restricción

Sitio de corte



extremos protruyentes (cohesivos)

Corte y empalme de moléculas de DNA (cut & paste)



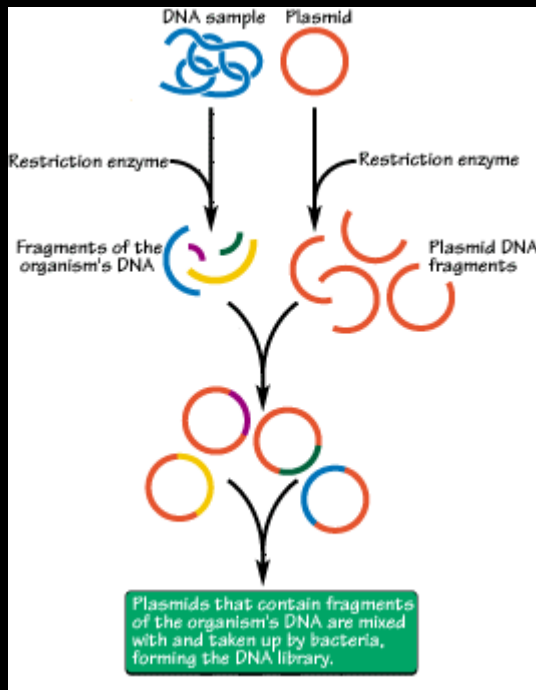
Enzimas de restricción

- Hay alrededor de 3600 enzimas de restricción bioquímicamente caracterizadas
- 588 están disponibles en forma comercial
- Cubren 211 de las 240 especificidades conocidas
- Base de datos de enzimas de restricción y sus sitios de reconocimiento y corte:
 - REBASE: <http://rebase.neb.com/>
 - Referencia: Nucleic Acids Research (2003) 31: 418-420

Manipulación *in vitro* del DNA

- Otras reacciones posibles:
 - Fosforilación/defosforilación (agregado/remoción de fosfatos en posición 5')
 - Ligación (unión covalente de dos moléculas de DNA a partir de un extremo 3' OH libre con uno 5' P libre)
 - Modificación de bases:
 - Metilación/demetilación (agregado/remoción de grupos metilo (CH₃) a una base)

Construcción de bibliotecas génicas



- Primer paso:
 - Obtener un número de plásmidos conteniendo cada uno un gen
- Segundo paso:
 - Introducir los plásmidos en bacterias para replicarlos (transformación)
- Tercer paso:
 - Obtener clones bacterianos: cada clon (colonia) contiene en teoría un gen foráneo

Hibridación

- El fenómeno de apareamiento de bases (A-T y C-G) para formar una doble hélice se llama hibridación, dado que puede utilizarse para formar DNA híbrido compuesto por cadenas de diferentes orígenes

CTGATGGTCATGAGCTGTCCGATCGATCAT

TACTCGACAGGCTAG

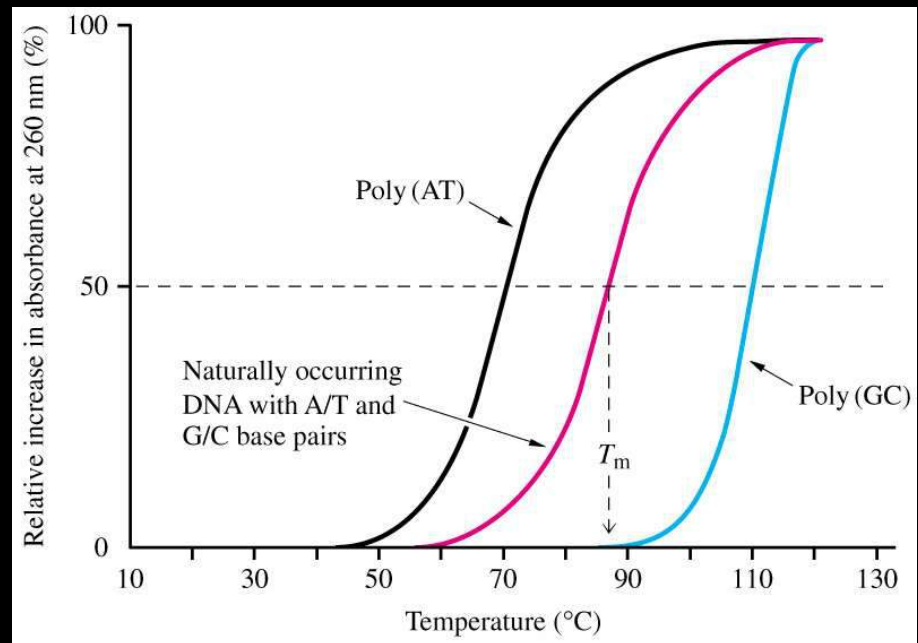
Hibridación

- El fenómeno de hibridación ocurre en solución, a un determinado pH, a una determinada temperatura, a una determinada concentración de sales, etc.
- Dependiendo de estas condiciones, es posible que en los híbridos existan malos apareamientos (A-C, por ejemplo)
- El origen de cada una de las hebras es irrelevante, sólo importa la secuencia (que un número significativo de bases sea complementario entre las dos)

DNA melting

Es posible disociar las dos hebras que forman una doble hélice de DNA aumentando la temperatura (por ejemplo)

La temperatura a la cual la mitad de la molécula de DNA (la mitad de los enlaces) está disociada, se conoce como T_m



Hibridación: usos

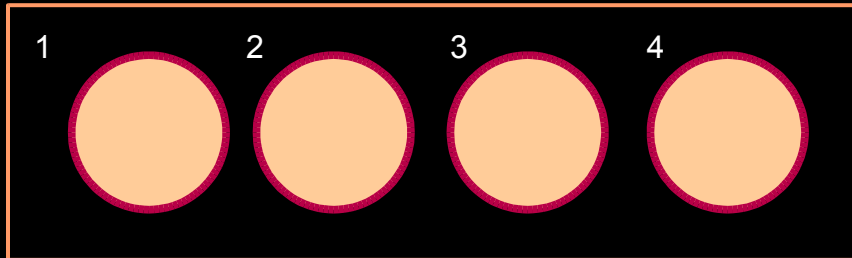
- Dado que dos hebras de DNA complementarias van a hibridar entre ellas si son puestas en las condiciones apropiadas, podemos aprovechar este comportamiento para obtener información sobre DNAs desconocidos
- Pequeños fragmentos de DNA conocidos pueden ser utilizados como sondas para testear la presencia de secuencias complementarias (en otros genomas, por ejemplo)
- Algunas aplicaciones de este concepto son:
 - Southern blots (DNA)
 - Northern blots (RNA)
 - *In situ* hybridization
 - Dot blots

Sondas de DNA

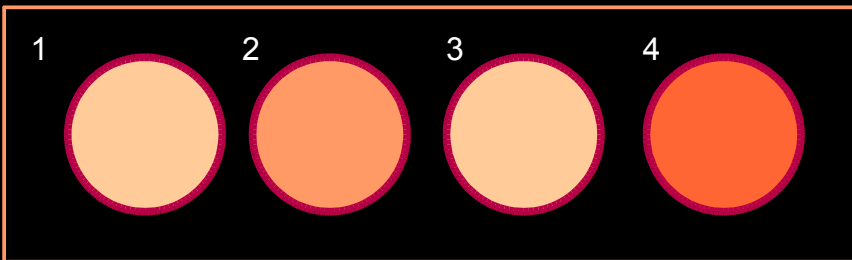
CTGATGGTCATGAGCTGTCCGATCGATCAT

|||||

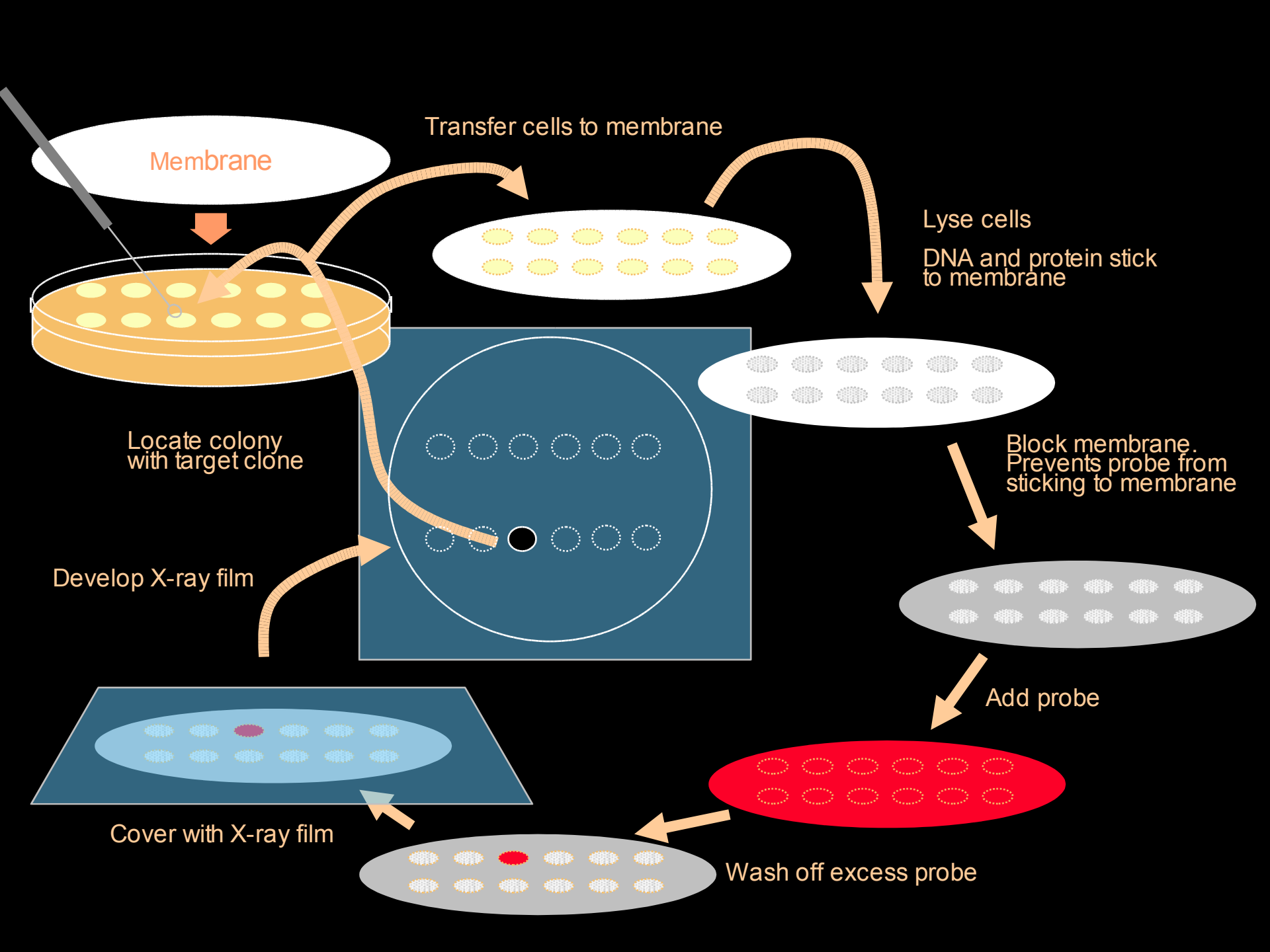
TACTCGACAGGCTAG



Antes de hibridar



Después de hibridar



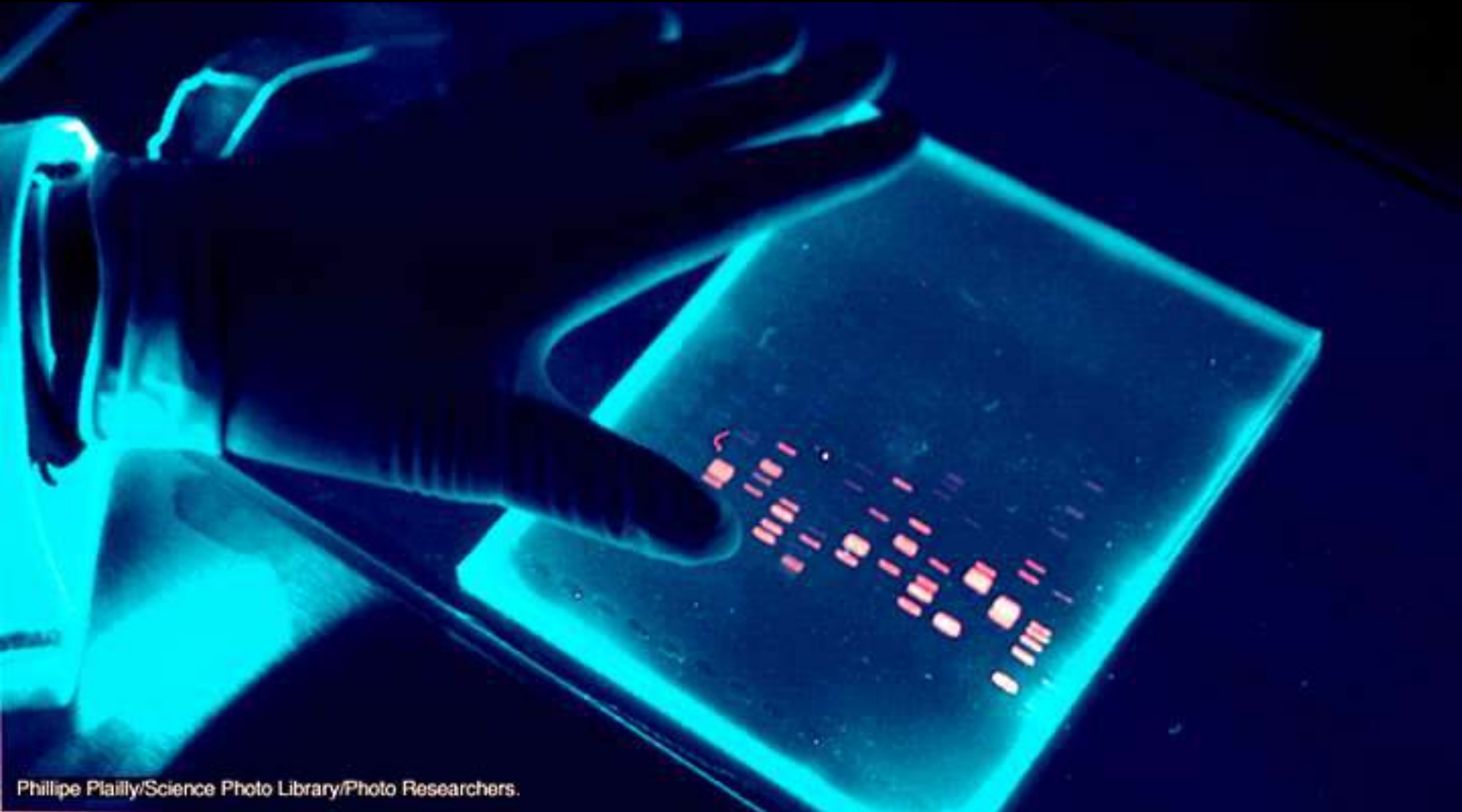
Separación de fragmentos de DNA: electroforesis

- La molécula de DNA tiene carga eléctrica
 - Los fosfatos proveen una carga neta negativa al DNA
 - Esta carga es homogénea a lo largo de toda la molécula
- En un campo eléctrico, el DNA migra hacia el polo positivo
- Si la migración se realiza en un medio semi-sólido (por ejemplo agar, una especie de gel) esta se dificulta, dado que el DNA debe pasar a través de poros en el gel.
- Como consecuencia de la necesidad de atravesar estos poros, las moléculas de DNA más grandes verán retrasada su movilidad, mientras que las más pequeñas migrarán una mayor distancia en el mismo tiempo.

Electroforesis, parte 2

- En un mismo “gel” es posible evaluar distintas mezclas de fragmentos (muestras)
- El DNA no es visible a simple vista
- Sin embargo utilizando bromuro de etidio (un colorante que se intercala entre las bases del DNA), es posible lograr que el DNA emita fluorescencia cuando es iluminado con luz ultravioleta

Electroforesis: foto



Southern blots

- Los fragmentos de DNA separados electroforéticamente pueden transferirse del gel a otros soportes (por ejemplo papel)
- Se utilizan papeles especiales (nitrocelulosa) u otro tipo de soportes (nylon) para lograr que el DNA quede adherido al soporte y poder realizar incubaciones, reacciones, etc sin peligro de que el DNA se “suelte”
- Esto, junto a la hibridación, son la base de la técnica conocida como “Southern Blot”

Southern blot esquemático

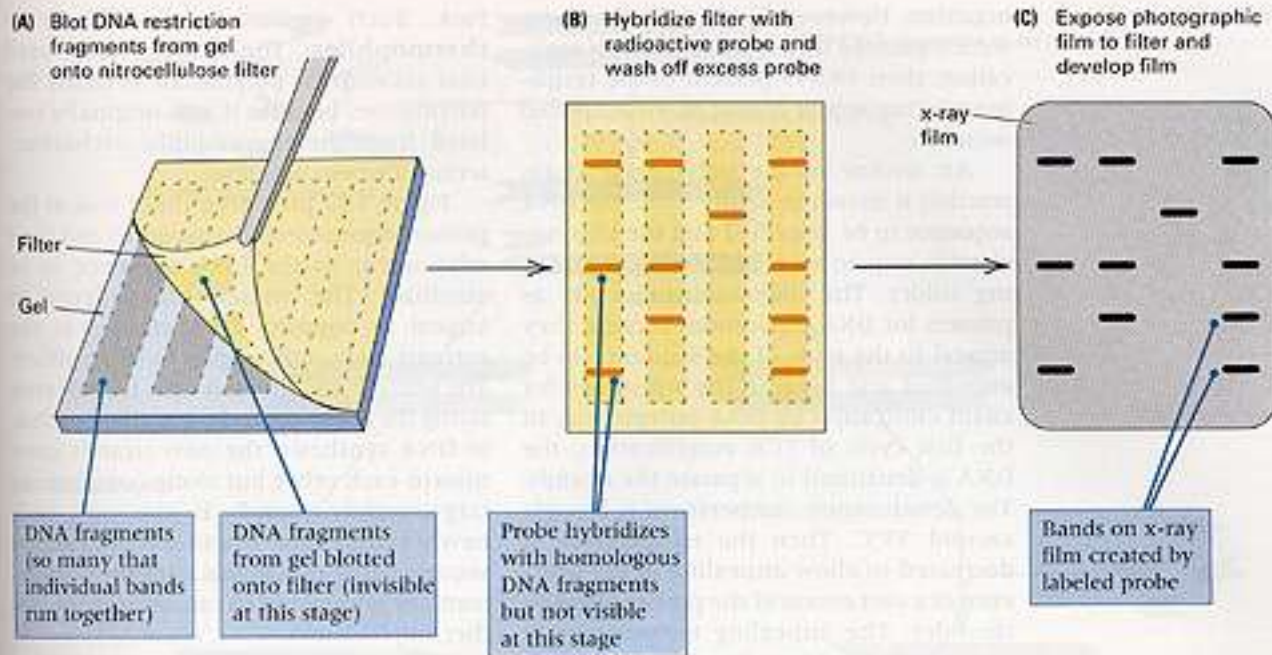


Figure 5.35

Southern blot. (A) DNA restriction fragments are separated by electrophoresis, blotted from the gel onto a nitrocellulose or nylon filter, and chemically attached by the use of ultraviolet light. (B) The strands are denatured and mixed with radioactive probe DNA, which binds with complementary sequences present on the filter. The bound probe remains, whereas unbound probe washes off. (C) Bound probe is revealed by darkening of photographic film placed over the filter. The positions of the bands indicate which restriction fragments contain DNA sequences homologous with those in the probe.

Un southern blot real

1 a 11: clones bacterianos conteniendo distintos genes de interés clonados

L: marcadores de tamaño

Izquierda: gel teñido con bromuro de etidio

Derecha: southern blot revelado con una sonda correspondiente a uno de los genes de interés

