



Guía N° 4: TRANSCRIPCIÓN DE ADN Ó SÍNTESIS DE ARN

El flujo de la información genética en las células normalmente es:

ADN → ARN → Proteína

La transcripción se define, básicamente, como la síntesis de ARN usando un ADN patrón o molde, mientras que la síntesis de proteína a partir de un ARN patrón es llamado **traducción**.

Clases de ARN

Existen tres clases de ARN:

- ARN mensajero (**mARN**)
- ARN de transferencia (**tARN**)
- ARN ribosomal (**rARN**)

En general, todos estos ARN son de hebra simple, pero el ARN de transferencia y ARN ribosomal contienen extensas regiones de doble hélice (la cadena de nucleótidos se dobla formando una horquilla o loop y stem loops). Los ARN más pequeños son los tARNs, que contienen alrededor de 75 nucleótidos, mientras que el más grande esta entre los mARN, que pueden tener más de 5.000 nucleótidos.

Enzima de la síntesis

Todos los ARN celulares son sintetizados por la **ARN polimerasa** de acuerdo a las instrucciones dadas por un ADN-patrón (Donde se está transcribiendo o molde). La dirección de la síntesis de ARN es 5'→3', similar a la síntesis de ADN.

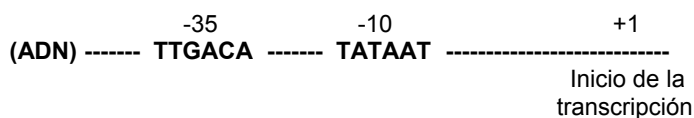
La actividad de la ARN polimerasa es diferente a la actividad de la **ADN polimerasa** porque no necesita un "primer" y no posee una actividad nucleasa. Otra diferencia es que el ADN patrón es totalmente conservado después de la síntesis de ARN.

La secuencia de ADN transcrito por la ARN polimerasa en ARN es llamado **UNIDAD DE TRANSCRIPCIÓN** y el producto de la síntesis (ARN) es el **TRANSCRIPTO PRIMARIO** o **pre ARN**.

LA ARN POLIMERASA DE E. COLI (PROCARIOTES):

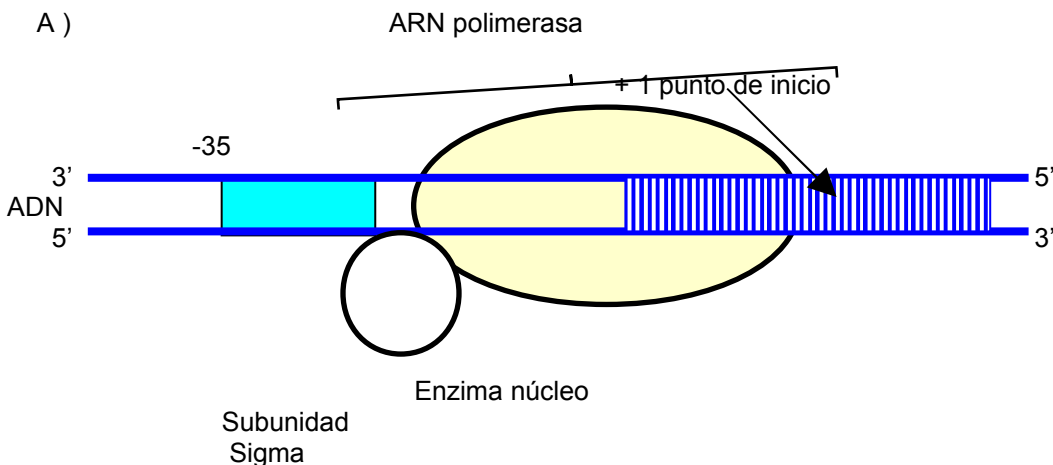
Formada por sub-unidades $\alpha_2\beta\beta'\sigma$. La sub-unidad **sigma (σ)** aumenta la velocidad y especificidad de la transcripción, debido que ayuda a la ARN polimerasa a reconocer la señal de inicio en el **ADN-patrón**. La sub-unidad sigma estaría reconociendo una región particular del ADN, que se encuentra a **35 bases** antes del sitio de inicio de la transcripción. Esta región del ADN recibe el nombre de región **PROMOTORA** y además de esta región, en E. coli existe otra región promotora a **10 bases** antes del inicio de la transcripción y está región del ADN esta relacionada con el lugar donde se une la enzima núcleo.

Región Promotora (E.coli - Procariontes)



El signo negativo indica que las bases se encuentran antes del sitio de inicio de la transcripción

Proceso de síntesis en E. Coli: La unión de la ARN polimerasa a estas regiones promotoras permiten un desenrollamiento local de 17 pares de bases de la molécula de ADN - patrón (llamada burbuja de transcripción). La síntesis de ARN siempre comienzan con una **base purina** (GTP ó ATP). La sub-unidad sigma se disocia cuando se han transcrito 12 bases. Esta subunidad una vez suelta puede unirse a otra enzima núcleo, permitiendo el inicio de una nueva síntesis.



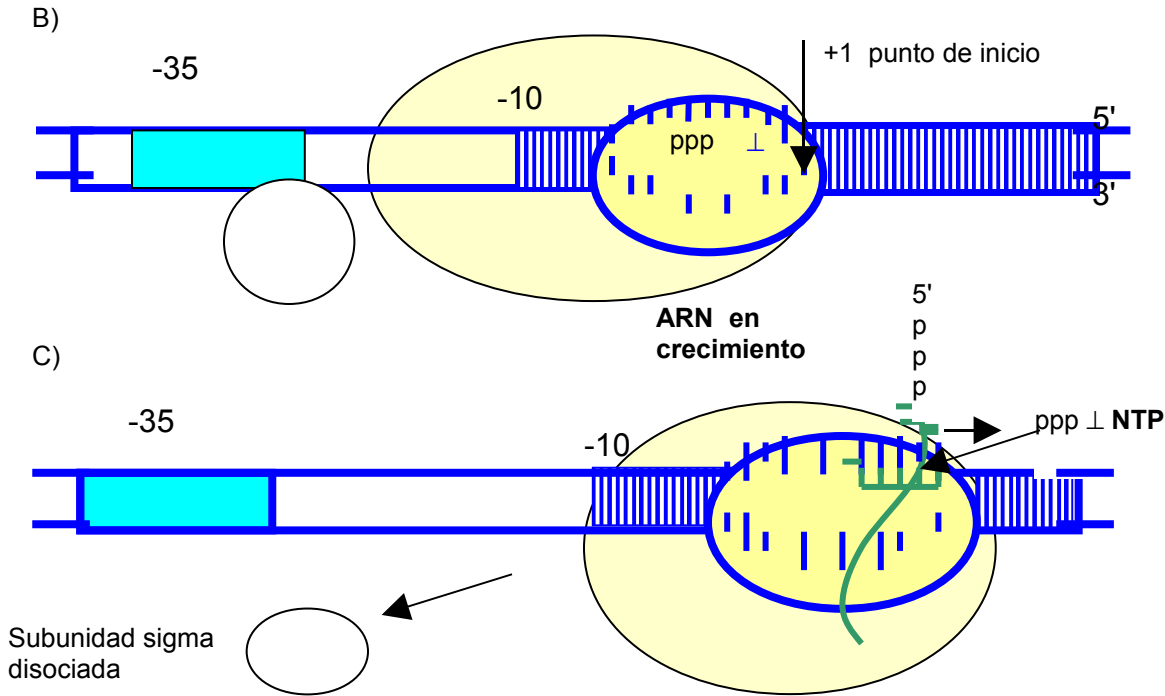


Figura 2. Inicio y elongación de la cadena de ARN (transcripción): (a) Unión de la holoenzima a la región promotora (complejo cerrado), (b) Separación de la doble hebra (complejo abierto) y (c) Elongación en dirección 5' → 3', a los 12 nucleótidos transcritos la subunidad sigma se disocia.

Terminación de la Transcripción.

La terminación de la transcripción es tan controlado como su iniciación. El ADN-patrón contiene señales de detención para la transcripción. Existen dos mecanismos de terminación en *E. coli*: uno es dependiente de una proteína auxiliar llamada factor rho y otro independiente de este factor.

Terminación independiente del factor rho. Esto es distinguido por la presencia de una secuencia rica GC en el ADN seguido por cinco o seis adeninas. EL ARN transcripto lógicamente también posee esta región rica en GC (es la complementaria), región que es capaz de formar un "loop" o estructura de horquilla como resultado de interacción entre bases complementarias. Además seguido de esta región rica en GC, en el extremo 3' del ARN aparece una secuencia rica en uracilo. Se ha observado que la unión Uracilo con Adenina de la hebra de ADN molde es menos estable y así el ARN es capaz de disociarse (ver fig. 3). Por consiguiente el duplex ADN-ARN se suelta y se libera el ARN sintetizado.

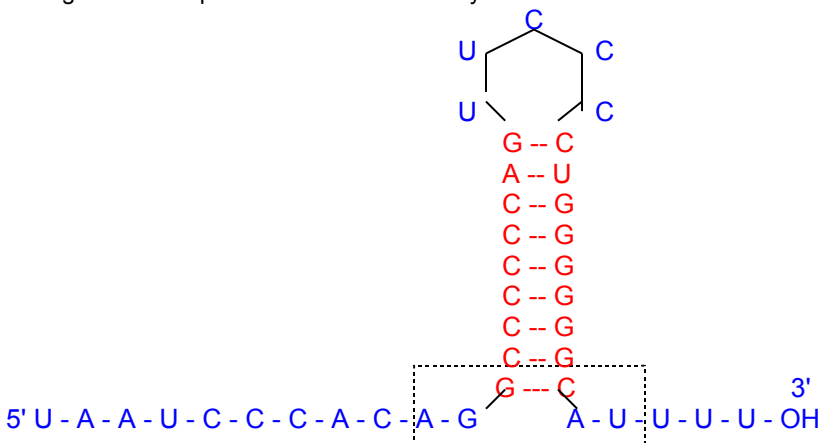


Figura 3. Estructura secundaria del ARN sintetizado (señal de termino).

Terminación dependiente del factor rho. También requiere la presencia de la estructura de horquilla (región rica en GC). Sin embargo, la *secuencia rica en U está ausente*. El factor rho, es una proteína que tiene una alta afinidad por ARN de hebra simple. Cuando este factor se une al ARN, hidroliza (rompe) ATP y la energía liberada es capaz de mover este complejo proteico (factor rho) a lo largo del ARN que se esta sintetizando en dirección a la burbuja de transcripción. Cuando llega a dicha burbuja, el factor rho disocia el híbrido ADN-ARN por un mecanismo desconocido, liberando el ARN al citoplasma (ver fig. 4).

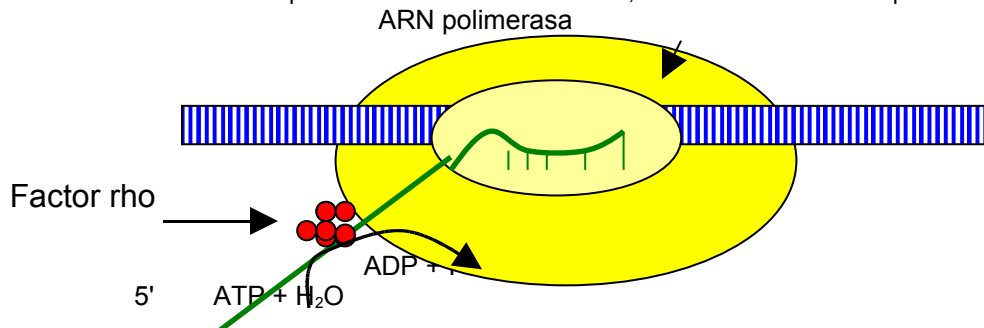


Figura 4. Terminación dependiente de rho

La transcripción puede ser bloqueada por **inhibidores específicos**, llamados comúnmente antibióticos. Por ejemplo, el antibiótico **rifampicina** inhibe la iniciación de la síntesis de ARN, mientras que la **actinomicina D** bloquea la elongación del ARN.

TRANSCRIPCIÓN EN EUCARIOTES:

La maquinaria transcripcional de eucariontes es lejos más complejo que en procariontes. Una importante consideración es que en eucariontes superiores solamente una pequeña proporción del genoma es expresada a ARN (cerca de 10% como máximo). Una considerable proporción del genoma de eucariontes existe permanentemente como Heterocromatina. Adicionalmente, hay proteínas accesorias específicas llamadas **FACTORES DE TRANSCRIPCIÓN** y son requeridos para una transcripción eficiente de todos los genes de eucariontes.

Diferente al factor bacteriano sigma (σ), *los factores de transcripción no se unen a la polimerasa directamente*. Ellos forman complejos con la cromatina antes del inicio de la transcripción y *actúa como un regulador positivo de la transcripción*.

Enzimas de la transcripción en eucariontes.

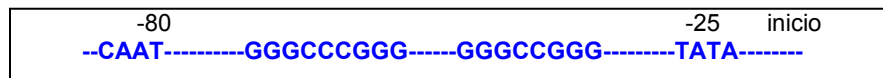
El núcleo de eucariontes contiene tres tipos de ARN polimerasas:

- **ARN polimerasa I** (localizada en el nucléolo) transcribe los genes de ARN ribosomal. El transcrito primario producido corresponde a un pre-rARN que posteriormente sufre algunas modificaciones para transformarse en una rARN maduro.
- **ARN polimerasa II** (se encuentra en el nucleoplasma) transcribe los genes que codifican para proteína y el transcrito primario corresponde a un ARN nuclear heterogéneo (**hnARN**), que son los precursores del ARN mensajero citoplasmático.
- **ARN polimerasa III** (nucleoplasmática) transcribe el rARN 5S (ARN ribosómico) y los tARNs (ARN de transferencia).

La reacción catalizada por todas las polimerasas en eucariontes es bioquímicamente la misma que la reacción de las enzimas de *E. coli* (procariontes).

Región promotora para la ARN polimerasa II.

Los genes transcritos por la polimerasa II producen los ARN mensajeros. Estos incluyen los genes que codifican a las proteínas que siempre son expresadas (**constitutivamente**) en todos los tejidos y aquellos genes que solamente son expresados en un tejido particular de un organismo multicelular o que esta regulado por la presencia o ausencia de un sustrato particular, hormona o estímulo ambiental. Los promotores para la ARN polimerasa II están localizados en el lado 5' del sitio de inicio de la transcripción.



Estas secuencias tienen la función de unir los factores de transcripción específicos en vez de dirigir los puntos de contacto con la ARN polimerasa II, como ocurre en procariontes. Varios promotores están cercanos al punto de inicio de la transcripción, son necesarios e insuficientes para que se realice una eficiente expresión de los genes por la enzima ARN polimerasa II.

Además existen otros tipos de secuencias en el ADN molde llamadas "**ENHANCER**" que no tienen actividad promotora por sí mismas, pero que pueden **estimular** la transcripción y pueden actuar sobre considerables distancias de varios kilobases del sitio de inicio de la transcripción. Una característica de los "enhancer" es que realizan su acción independiente de la orientación que ellos tengan (Fig. 6)



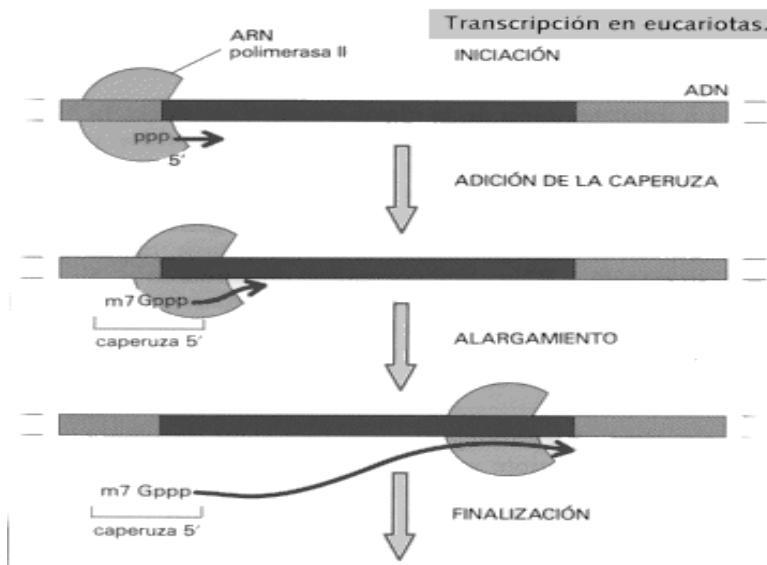
Figura 6. Esquema de la posición del "Enhancer" con respecto al promotor.

El transcrito primario de la ARN polimerasa II, son conocidas colectivamente como ARNs nucleares heterogéneos. Ellas pueden llegar a ser grandes moléculas (sobre 10 kilobases) debido que contiene aún los intrones (secuencia que **no** son traducidas a proteína) en la secuencia de ARN, pero ellas no estarán presente en el ARN maduro.

SINTESIS DE ARN MENSAJEROS

La síntesis de un mRNA en eucariontes:

- Comienza la transcripción de un ADN cuando una ARN polimerasa se encuentra a 20 o 30 nucleótidos después de la secuencia **TATA**. Cuando el transcrito (ó ARN) ha alcanzado 30 nucleótidos de largo, en su extremo 5' es colocado una **guanosina metilada** (*cap* – *caperuza 5'*)



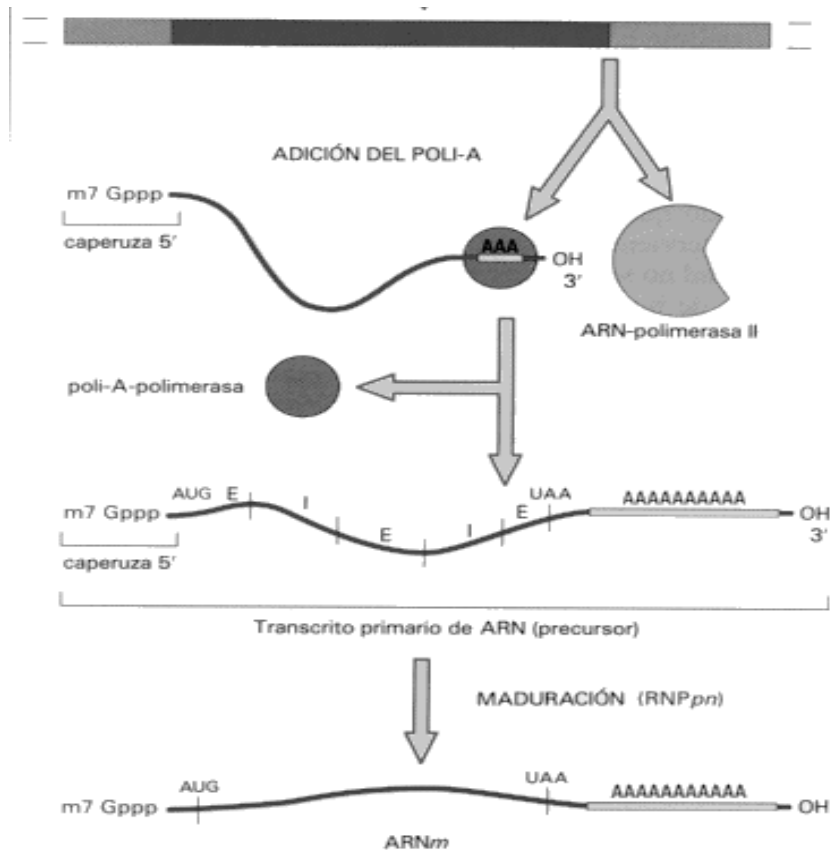


Fig. 7

- La ARN polimerasa se mueve a lo largo del ADN, transcribiendo tanto exones como intrones. Hasta que se transcribe la secuencia **AAUAAA**, después de cerca de 20 nucleótidos de esta secuencia, el transcrito es cortado; la reacción probablemente involucra una pequeña **ribonucleoproteína nuclear** (snRNP) que contiene una molécula de ARN ("U1 ARN") rico en uracilo. Una enzima adiciona una secuencia de 150 a 200 adeninas en el extremo 3' de la hebra de ARN naciente. El transcrito es procesado y son cortados los intrones, probablemente con la ayuda de otras snRNPs. Los intrones son doblados y cortados y los exones son unidos para formar finalmente el ARN mensajero maduro

La mayoría de los genes en eucariontes superiores son discontinuos. Las secuencias que se codifican en estos genes (**exones**) son separadas por secuencias que no son traducidas (**intrones**), que son removidos en la conversión del transcrito primario a mRNA maduro.

Modificaciones Post Transcripcionales: Muchos ARNs son modificados químicamente (Ej. metilados) y escindidos o cortados después de la transcripción. Los mARNs son modificados por acoplamiento de una larga cadena de poli-A a su extremo 3'. Los rARNs se generan a partir de una larga cadena precursora que finalmente es escindida para rendir los ARNs maduros.

CÓDIGO GENÉTICO

La secuencia de bases de un gen es colineal con la secuencia aminoacídica de su producto polipeptídico. El código genético es la relación entre la secuencia de bases en el ADN (o su ARN transcrito) y la secuencia de aminoácidos en una proteína.

Los aminoácidos son codificados por un grupo de tres bases (llamadas **codones**) comenzando de un punto determinado (ver fig. 8). Se pueden definir cuatro características generales del código genético:

- 1) El código no requiere de puntuación ni señal alguna para indicar el final de un triplete y el comienzo del siguiente, es decir el código genético no tiene "comas". Únicamente requiere de una molécula de mRNA que este ubicada correctamente para su lectura (5'→3') además de contar con el codón de inicio, **AUG** (en procariontes codifica la incorporación de N-formilmetionina y en eucariontes metionina).
- 2) 61 de 64 codones especifican algún aminoácido en particular. De esta forma, para la mayoría de los aminoácidos hay más de un triplete (o codón). En otras palabras, el código es **DEGENERADO**. Codones que especifican el mismo aminoácido son llamados **sinónimos**. La mayoría de los sinónimos difieren solamente en la última base del triplete. Esto presenta una ventaja, debido a que la mayoría de las mutaciones conducen a la formación de tripletes sinónimos no provocando un cambio en la secuencia peptídica.
- 3) La tercera base de los tripletes es menos específica que las dos primeras. La degeneración implica solamente la tercera base en la mayoría de los casos (excepto para Arginina, Leucina y Serina).
- 4) 3 de los 64 tripletes no codifican para aminoácido alguno y estos son **UAG, UAA y UGA** llamados originalmente codones "sin sentido o stop" **que constituyen las señales de termino de la síntesis de la cadena polipeptídica.**

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

Figura 8. Código genético.

SÍNTESIS DE PROTEÍNAS

La síntesis de proteínas tienen lugar en la superficie de los ribosomas. Dicha síntesis requiere que los aminoácidos sean “**activados**” en el citoplasma por la acción de **aminoacil- tARN-sintetasas**, que catalizan la formación de ésteres de aminoácidos de tARN homólogos (unión de un aminoácido con su tRNA respectivo) .

El aminoácido se une al extremo 3' del tARN que posee la secuencia **CCA**, mientras que el anticodón está a 80 Å de distancia en el otro extremo (ver fig. 9).

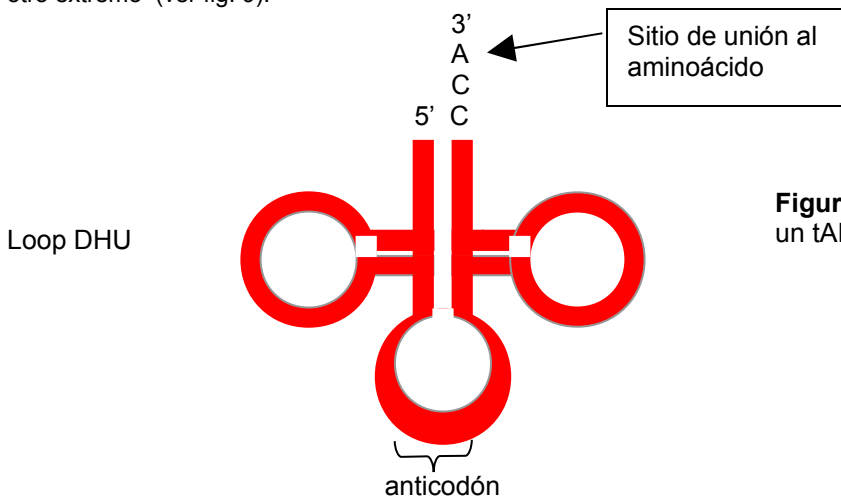


Figura 9. Estructura de hoja de trébol característica de un tARN

El ARN mensajero reconoce el anticodón de un tARN en vez de reconocer el aminoácido. Un codón en mRNA se aparea con el anticodón en el tARN. Algunos tARNs reconocen más de un codón porque la tercera base que se aparea de un codón es menos específica que las otras dos (ver código genético).

Hay tres fases en la síntesis de proteína: **inicio**, **elongación** y **termino**.

Inicio: En **procariontes** el ARN mensajero, la formilmetionina- tARN y la sub-unidad ribosomal 30 S, forman el complejo iniciación de la síntesis (en cambio para eucariontes este complejo está formado por mRNA, metionina-tARN y la sub-unidad 40 S).

La señal de inicio en el mRNA es **AUG** y esta precedida por una secuencia rica en purina que puede aparearse con el rARN 16 S de la sub-unidad 30 S del ribosoma. La subunidad mayor ribosomal de 50 S se une a este complejo para formar un **complejo de iniciación de 70 S**, que está listo para la siguiente fase (elongación).

Elongación. El ciclo de elongación consiste en (ver fig. 10) :

- 1) La unión de aminoacil-tARN (reconocimiento del codón).
- 2) Formación del enlace peptídico.
- 3) Transposición.

La dirección de lectura del mRNA es de 5' → 3' y el crecimiento de la cadena polipeptídica es hecho desde el extremo amino al extremo carboxilo.

Figura 3. Traducción del ARNm

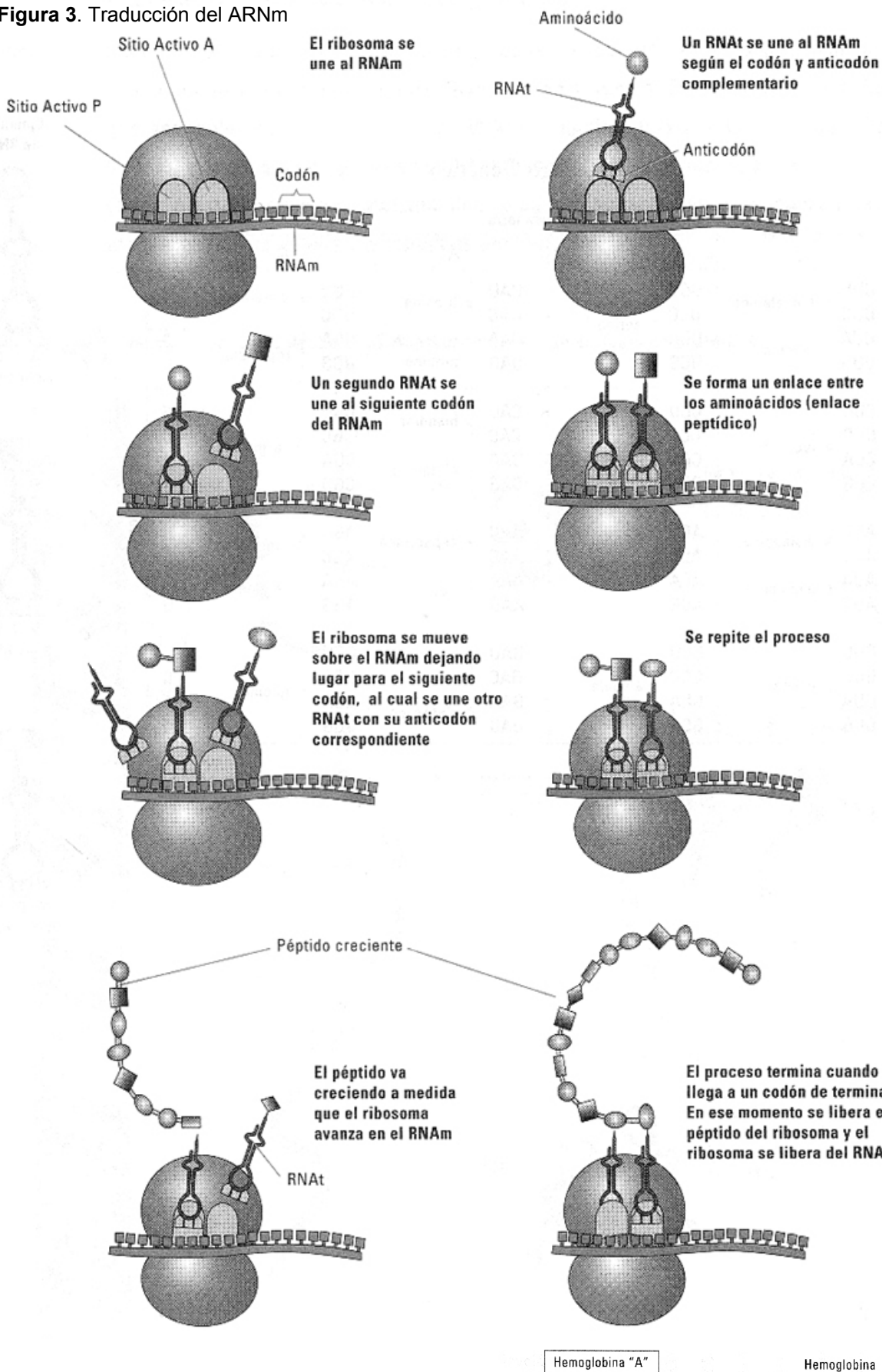


Figura 10. Fase de elongación de la cadena

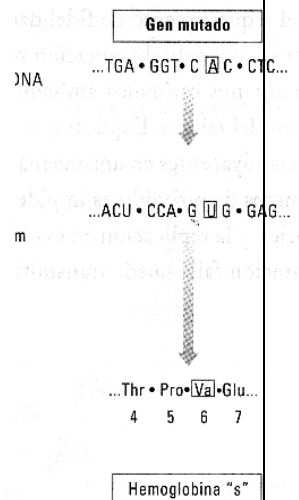
Termino.

La síntesis de proteína es terminada por factores de

La energía requerida tanto en la formación del complejo de iniciación son 70

Varios pasos en la síntesis de proteína son específicamente inhibida por toxinas y antibióticos. Por ejemplo, puromicina inhibe la síntesis de la cadena peptídica por su interacción con el peptidil - tARN en la ribosoma,

También existe síntesis proteica en las mitocondrias y en los cloroplastos,



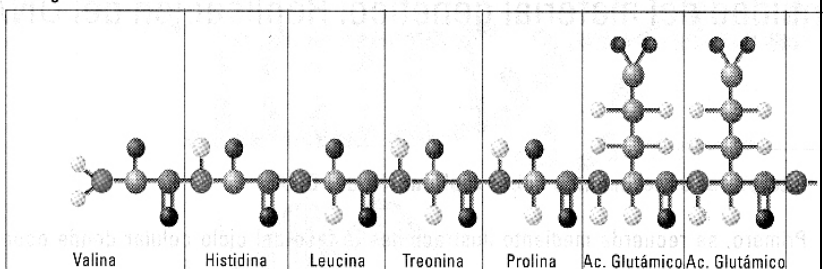
Alteraciones en la lectura

Tal como se señaló, el modelo de ADN de Watson y Crick sugiere que se forma una copia exacta de esta molécula cada vez que se autoduplica. Sin embargo, debido a "accidentes moleculares" se producen "errores" en el mensaje genético con su correspondiente expresión anormal en el fenotipo.

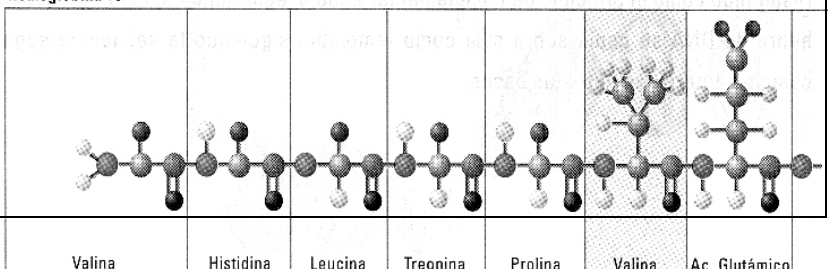
Las variaciones en el mensaje hereditario llevado por el ADN reciben el nombre de mutaciones; éstas pueden aparecer debido a cambios en un gen -segmento de la molécula de ADN- o a cambios en el número o en la estructura de los cromosomas. Se habla entonces de mutación génica y mutación cromosómica.

Un gen puede resultar alterado por una reordenación accidental de las bases nitrogenadas que

Hemoglobina normal



Hemoglobina "s"





constituyen el mensaje. Una analogía puede explicar esto: la palabra "tren" tiene un significado claro para todo el mundo de habla hispana. Cuando tú vez esta palabra la interpretas correctamente. Supón ahora que se produce una reordenación de letras que conforman esta palabra de modo que resulta "en rt". Claramente, esto no tiene ningún significado y no podemos interpretarlo. La mutación génica también puede resultar por sustracción o adición de una base del código. Volviendo a nuestro ejemplo, si a la palabra "tren" se quita la letra "t" o se le agrega la letra "a" resultan palabras carentes de significado.

Lo mismo sucede con el código genético; la reordenación de las bases nitrogenadas, la pérdida o ganancia de un nucleótidos, la pérdida o ganancia de un triplete de bases alteran el código genético y la célula es incapaz de interpretar el mensaje alterado. Esto significa que no se producirá la o las proteínas de acción enzimática correspondientes, por lo cual la secuencia de reacciones bioquímicas no se producirán, resultando un individuo anormal.

Una cadena "mutante" del ADN puede diferenciarse de la cadena normal por un sólo nucleótido. Sin embargo, este pequeño cambio en el mensaje hereditario tiene un efecto en la célula. Podría causar sólo una pequeña variación en la estructura de una teína como una enzima. Como resultado de este cambio podría verse afectada la actividad o la eficacia de esta enzima en la reacción que cataliza.

Trabajo práctico: Síntesis de un péptido - la insulina

A partir de las instrucciones entregadas por el profesor acerca del plegamiento de proteínas tras la síntesis de la cadena aminoacídica y considerando la información contenida en esta guía, deberás obtener la estructura de una molécula de INSULINA (hormona relacionada con la regulación de la glicemia) a partir de la secuencia simplificada de su gen.

I. Gen de la insulina (ubicado en el cromosoma nº 11)

- Se han omitido las bases que originan tripletes de inicio y final
- Los espacios disponibles bajo la secuencia son para anotar el código del ARN_m correspondiente (línea superior), así como el orden de los aminoácidos (línea inferior)

AAGCAGTTGGTTGTGGATACGCCGAGCGTGGATCAGCTCCGTGATATGGATCAGACGCCGCTCGCCCCG

AAGAAGATGTGTGGTTTCCGTTGTGCCGCCCTCCGTCTCTGGATGTTTCAGCCGGTTCAGCTCGATCCG

CCGCCGGTTCGTCCGAGCGATGTTGGTGATCGTGATCTCCCGAGCGATGTTTTGCCCGTATCAGCTC

GTTACGACGTGTAGCTAGACGAGCGATATGGTTGATC TCTTGATGACGTTG

II. Especificaciones de terminación y plegamiento:

- a) Número de cadenas peptídicas independientes = 3, de las cuales sólo dos terminan formando la insulina.
- Cadena B: Fenilalanina 1 hasta Alanina 30
 - Cadena C: Treonina 1 hasta Arginina 35 (cadena que es cortada y perdida)
 - Cadena A: Glicina 1 hasta Asparagina 21
- b) Puentes disulfuro (entre cisteínas):
- Cys6 - Cys11 en la cadena A
 - Cys7 cadena A con Cys7 cadena B
 - Cys20 cadena A con Cys19 cadena B
- c) Número de segmentos que forman alfa-hélice = 3
- Dos segmentos en la cadena A entre los aminoácidos Gly1 - Ile10, Ser12 - Glu17
 - Un segmento en la cadena B entre los aminoácidos Ser 9 - Gly20
 - Tener presente que los aas hidrofóbicos (Gly, Ala, Val, Leu, Ile, Trp, Pro, Met, Phe) se orientan hacia el interior de la proteína, mientras que los hidrofílicos (Asn, Gln, Tyr, Cys, Ser, Thr) lo hacen hacia fuera

III. Especificaciones de diseño:

- Los aminoácidos pueden representarse como círculos u otro dibujo simple.
- Cada cadena (A y B) debe representarse de un color distinto
- Los 3 puentes disulfuro deben quedar destacados y las secuencias que forman hélices alfa pueden representarse como un espiral simple
- Es preferible realizar varios borradores hasta diseñar un modelo que cumpla con todas las especificaciones.
- Basándose en el diseño REVISADO Y APROBADO por el profesor, construir un modelo 3D de la insulina, utilizando alambre, plastilina y/o pelotitas de plumavit.

Anexo: Clasificación y abreviaturas de los 20 aminoácidos:

Polares	Asparagina: Asn	Glutamina: Gln	Tirosina: Tyr	Serina: Ser	Treonina: Thr
Con carga eléctrica	Ácido aspártico: Asp	Ácido glutámico: Glu	Arginina: Arg	Lisina: Lys	Histidina: His
	Glicina: Gly	Alanina: Ala	Valina: Val	Leucina: Leu	Isoleucina: Ile
No polares	Triptófano: Trp	Prolina: Pro	Cisteína: Cys	Metionina: Met	Fenilalanina: Phe

EJERCICIOS

- 1.- Escribir la secuencia complementaria en dirección 5' →3'
- a) 5'- GATCAA- 3'
- b) 3'- ACTGGCCTAA -5'

- c) 5'- ACCTAGGGTA -5'
- d) 3'- CCTGGATTAAG -5'



2.- Compare la ADN polimerasa I y la ARN polimerasa de E.coli en cuanto a:

- a) estructura de subunidades.
- b) precursores activados
- c) dirección de elongación de la cadena
- d) actividad exonucleasae) conservación del ADN molde
- f) necesidad de una hebra particular.
- d) energética de la reacción de elongación

3.- Escribir las secuencias de los mRNA sintetizados de los siguientes ADN

- a) 3'- ATTGCCATGCTA-5'
- b) 5'- AGCTACTTAACG-3'
- c) 3'-GGACCTACGTT.5'

Además indique cuales son los codones sin sentido presentes.

4.- Cuantos aminoácidos pueden ser codificados de las siguiente secuencia de mRNA.

- a) Asumiendo que comienza la lectura en el extremo 5' inmediatamente.
 - b) Comenzando en forma normal establecida por el código genético.
4. 1- 5'- UUGCCU AUGGAUUGGAUG-3'
4. 2- 3'-AUGUAGGUAAAAGGUAGGCC-5'
4. 3- 3'- AGCGCCAGUGACCAUGUA-5'
5.- Que secuencia es formado por la adición de un poli (UUAC) a un sistema de síntesis de proteínas.

6.- El ARN transcrito de una región del ADN del fago G4 conteniendo la siguiente secuencia (este tipo de ADN tiene una traducción superpuesta)

5' - TCCTCATTT - 3'

Esta región codifica para diferentes proteínas. ¿Cuales son sus secuencias?.

Mutaciones:

Las mutaciones son causadas por cambios en la secuencia de bases de ADN. Los principales tipos son sustitución, supresión e inserción. La sustitución de un par de bases por otra es la más común mutación. Una transición es el reemplazo de una purina por otra purina (lo mismo ocurre con las pirimidinas). Una transversión es el reemplazo de una purina por una pirimidina o viceversa. Muchos potenciales cancerígenos pueden ser detectados por su acción mutagénica en bacterias.

Gen Salvaje 5'- CGACTGT-3'
3'- GCTGACA-5'

Transición (cambio del par A-T por G-C)
5'- CG**G**CTGT-3'
3'- G**C**C**G**ACA-5'

Transversión (cambio del par A-T por T-A)
5'- CGTCTGT-3'
3'- GCA**G**ACA-5'

Inserción (colocar un par extra ej: G-C)
5'- CGA**G**CTGT-3'
3'- GCT**C**GACA-5'

Supresión (suprimir el par A-T)
5'- CGCTGT-3'
3'- GCGACA-5'