

TRASCRIZIONE E TRADUZIONE PT1

La **trascrizione** e la **traduzione** sono i mezzi mediante i quali le cellule leggono, o esprimono, le istruzioni genetiche presenti nei loro geni. Poiché dallo stesso gene si possono ottenere molte copie identiche di RNA, e ciascuna molecola di RNA può dirigere la sintesi di molte molecole proteiche identiche, le cellule possono sintetizzare una grande quantità di proteine rapidamente quando è necessario. Ma ciascun gene può anche essere trascritto e tradotto con un'efficienza diversa, permettendo alla cellula di produrre grandi quantità di alcune proteine e piccolissime di altre.

Le varie tappe della sintesi proteica: gli amminoacidi da utilizzare per la formazione della catena polipeptidica vengono attivati dall'ATP e da un enzima (specifico per ogni amminoacido) contenuto nel citoplasma cellulare: l'amminoacido attivato viene quindi trasferito sul RNA transfer (tRNA). Esisterebbe una forma specifica di tRNA per ogni amminoacido; questo viene trasferito in corrispondenza dei ribosomi su una catena peptidica in formazione, rispettando la sequenza determinata dal codice iscritto nella memoria del DNA nucleare. L'informazione genetica viene trasferita dal DNA alle sedi dove si svolge la sintesi delle proteine, tramite il RNA messaggero. Tale meccanismo richiede la sintesi nucleare del RNA messaggero, nel quale viene riprodotta l'informazione genetica. A tale scopo interviene uno speciale enzima, detto RNA polimerasi-DNA-dipendente, che, utilizzando il DNA come modello, sintetizza mRNA in presenza di ATP. La formazione nucleare di mRNA presenta analogie con il meccanismo di duplicazione del DNA: si forma infatti una singola catena di RNA nel quale la sequenza delle basi azotate è complementare a quella di una delle due eliche del DNA. In questo modo l'informazione contenuta nella sequenza delle basi del DNA viene trascritta nella sequenza, a essa complementare, delle basi del mRNA sintetizzato: questo processo è quello di **trascrizione**. Un singolo nucleotide del DNA non può essere responsabile della posizione nella catena polipeptidica di un singolo corrispondente amminoacido, dato che gli amminoacidi di norma utilizzati per le sintesi delle proteine sono circa una ventina mentre le basi azotate costitutive degli acidi nucleici sono soltanto quattro. Il codice deve perciò essere costituito da una sequenza di più nucleotidi. L'associazione a tre a tre delle basi azotate dà luogo a 64 diverse combinazioni, si è avanzata l'ipotesi che il codice fosse formato da triplette di basi nucleotidiche. Alcuni amminoacidi possono essere codificati da più di una tripletta (fenomeno della degenerazione del codice), mentre alcune triplette non hanno significato ai fini della sintesi proteica (triplette "non senso"), ma significano **"fine della catena polipeptidica"**. In molti casi il significato delle triplette dipende da due sole basi, mentre la terza ha scarsa o nulla significatività. Il modo con cui l'mRNA sintetizzato nel nucleo può raggiungere i ribosomi senza essere distrutto dalle ribonucleasi, cioè un gruppo di enzimi che presiedono all'idrolisi degli acidi ribonucleici o RNA che abbondano nel citoplasma. Si ritiene che durante il trasferimento al citoplasma il mRNA sia rivestito di particolari proteine che, oltre a proteggere come in un involucro la molecola dell'acido nucleico, potrebbero anche regolare la combinazione del mRNA ai ribosomi. Il secondo processo fondamentale della sintesi proteica è la **traduzione**.

La traduzione è la formazione di proteine sullo stampo di mRNA. Con questo processo le triplette di nucleotidi sul mRNA vengono decifrate come i corrispondenti amminoacidi individuali. L'mRNA arrivato nel citoplasma si dispone su corpuscoli detti **ribosomi**, che si trovano nel reticolo endoplasmico rugoso. A esso si associano molecole di **transfer RNA (tRNA)**, che portano legate in modo specifico un solo amminoacido ciascuno. Come amminoacil-tRNA, questo si sistema sul ribosoma in corrispondenza della tripletta (codone) che rappresenta il suo specifico amminoacido. È in grado di riconoscerla grazie a una sequenza nucleotidica, detta **anticodone**, complementare al codone dell'amminoacido che trasporta. Man mano che il mRNA scorre sul ribosoma, i vari tRNA si mettono in posizione e si attaccano tra loro con legami peptidici, formando alla fine la proteina codificata dal quel particolare mRNA.

Il primo passaggio della lettura di una parte necessaria delle istruzioni genetiche di una cellula è quello di copiare una porzione particolare della sequenza nucleotidica del suo DNA (un gene) in una sequenza nucleotidica di RNA. L'informazione nell'RNA, anche se copiata in un'altra forma chimica, è ancora scritta essenzialmente nello stesso linguaggio del DNA, il linguaggio di una sequenza nucleotidica. Da qui il nome **trascrizione**.

Come il DNA, l'RNA è un polimero lineare composto da quattro tipi di subunità nucleotidiche unite insieme da legami fosfodiesteri che differisce chimicamente dal DNA per due aspetti:

1. I nucleotidi nell'RNA sono ribonucleotidi, cioè contengono lo zucchero ribosio invece del deossiribosio;
2. Sebbene, come il DNA, l'RNA contenga le basi adenina, guanina e citosina, contiene la base uracile invece della timina del DNA. Poiché U, come T, può appaiarsi mediante legami idrogeno con A, le proprietà di appaiamento complementare descritte per il DNA si applicano anche all'RNA. Nell'RNA si trovano anche altri tipi di coppie di basi: per esempio, occasionalmente G si appaia con U.

Nonostante queste differenze chimiche siano piccole, DNA e RNA differiscono in modo notevole nella loro struttura globale. Mentre il DNA si trova sempre nelle cellule come elica a doppio filamento, l'RNA è a singolo filamento. Le catene di RNA si possono perciò ripiegare in una forma particolare, proprio come una catena polipeptidica si ripiega prendendo la forma finale di una proteina. La capacità di ripiegarsi in forme tridimensionali complesse permette ad alcune molecole di RNA di avere funzioni strutturali e catalitiche precise.

Tutto l'RNA in una cellula è prodotto dalla trascrizione del DNA, un processo che ha alcune somiglianze con il processo della replicazione del DNA. La trascrizione inizia con lo svolgimento di una piccola porzione della doppia elica del DNA per esporre le basi su ciascun filamento di DNA. Uno dei due filamenti della doppia elica di DNA agisce quindi da stampo per la sintesi di una molecola di RNA. Come nella replicazione del DNA, la sequenza nucleotidica della catena di RNA è determinata dall'accoppiamento complementare delle basi fra i nucleotidi in arrivo e lo stampo di DNA. Quando c'è una buona corrispondenza il ribonucleotide in arrivo viene legato covalentemente alla catena di RNA in crescita in una reazione catalizzata enzimaticamente. La catena di RNA prodotta dalla trascrizione, viene perciò allungata un nucleotide alla volta e ha una sequenza nucleotidica che è esattamente complementare al filamento DNA usato come stampo.

La trascrizione, tuttavia, differisce dalla replicazione del DNA per diversi aspetti. A differenza di un filamento di DNA appena formato, il filamento di RNA non rimane legato da legami idrogeno al filamento stampo di DNA ma, appena dietro la regione in cui i ribonucleotidi vengono aggiunti, la catena di RNA è spostata e l'elica di DNA si riforma. Così le molecole di RNA prodotte dalla trascrizione sono rilasciate dallo stampo di DNA come singoli filamenti. Inoltre, poiché sono copiate soltanto da una regione limitata del DNA, le molecole di RNA sono molto più brevi delle molecole di DNA. Gli enzimi che eseguono la trascrizione si chiamano **RNA polimerasi**. Come la DNA polimerasi che catalizza la replicazione del DNA, le RNA polimerasi catalizzano la formazione dei legami fosfodiesteri che uniscono i nucleotidi per formare una catena lineare. L'RNA polimerasi si muove un passo alla volta lungo il DNA, svolgendo l'elica di DNA appena davanti al sito attivo di polimerizzazione per esporre una nuova regione del filamento stampo per l'accoppiamento complementare delle basi. In questo modo la catena di RNA in crescita è estesa di un nucleotide alla volta nella direzione 5'-3'. I substrati sono nucleosidi trifosfati (ATP, CTP, UTP e GTP); come per la replicazione del DNA, un'idrolisi di legami ad alta energia fornisce l'energia necessaria per spingere in avanti la reazione. Il rilascio qui immediato del filamento di RNA dal DNA durante la sua sintesi significa che molte copie di A possono essere prodotte dallo stesso gene in un tempo relativamente breve, in quanto la sintesi di altre molecole di RNA inizia prima che il primo RNA sia completato. Quando le molecole di RNA polimerasi si susseguono in questo modo una vicina all'altra, muovendosi a circa 20 nucleotidi al secondo, in un'ora da un singolo gene si possono sintetizzare più di mille trascritti. Sebbene l'RNA polimerasi catalizzi essenzialmente la stessa reazione chimica della DNA polimerasi, ci sono alcune differenze importanti fra le attività dei due enzimi. La prima è che l'RNA polimerasi catalizza l'unione di ribonucleotidi e non deossiribonucleotidi. In secondo luogo, a differenza delle DNA polimerasi coinvolte nella replicazione del DNA, le RNA polimerasi possono dare inizio ad una catena di RNA senza un primer.

Questa differenza può esistere perché la trascrizione non ha bisogno di essere così accurata come la replicazione del DNA. A differenza del DNA, l'RNA non conserva permanentemente l'informazione genetica nelle cellule. Le RNA polimerasi compiono circa un errore ogni 10^4 nucleotidi copiati in RNA (DNA 10^7) e le conseguenze di un errore nella trascrizione dell'RNA sono molto meno importanti che nella replicazione del DNA. Se un ribonucleotide non corretto viene aggiunto alla catena di RNA in crescita, la polimerasi può tornare indietro e il sito attivo dell'enzima può seguire una reazione di escissione che assomiglia all'inverso della reazione di polimerizzazione, eccetto che si usa acqua invece di pirofosfato e viene rilasciato un nucleoside monofosfato.

La trascrizione è un processo che trasferisce il messaggio contenuto nel DNA su un messaggero, mRNA, per destinare questo messaggio al processo di traduzione. Quindi la sequenza di DNA viene riprodotta su un messaggero di RNA, detto mRNA, che può uscire dal nucleo e tradurre il proprio messaggio in proteina. La maggioranza dei geni presenti nel DNA di una cellula specifica la sequenza degli amminoacidi di proteine; le molecole di RNA che sono copiate da questi geni sono gli **RNA messaggeri (mRNA)**. I **piccoli RNA nucleari (snRNA)** dirigono lo splicing del pre-mRNA per formare mRNA, che l'RNA **ribosomiale (rRNA)** forma il nucleo dei ribosomi e che gli **RNA transfer (tRNA)** formano gli adattatori che scelgono gli amminoacidi e li tengono in posizione su un ribosoma per incorporarli nelle proteine. Ciascun segmento trascritto di DNA si chiama **unità di trascrizione**. Negli eucarioti una unità di trascrizione di norma porta l'informazione di un singolo gene e perciò codifica o per una singola molecola di RNA o per una singola proteina (o un gruppo di proteine correlate se il trascritto iniziale di RNA viene sottoposto a splicing in più di un modo per produrre mRNA diversi).

Per trascrivere un gene la RNA polimerasi deve riconoscere dove iniziare sul genoma e dove finire. Il modo in cui le RNA polimerasi svolgono questi compiti è diverso nei batteri e negli eucarioti.

L'inizio della trascrizione è un passaggio particolarmente importante nell'espressione genica perché è il punto principale in cui la cellula regola quali proteine debbano essere prodotte e a che ritmo. Il nucleo dell'enzima della RNA polimerasi batterica è un complesso multisubunità che sintetizza RNA usando come guida uno stampo di DNA. Una subunità che si può staccare, chiamata **fattore sigma σ** , si associa al nucleo dell'enzima e lo assiste nella lettura dei segnali nel DNA che dicono dove iniziare a trascrivere. Insieme al fattore σ e il nucleo dell'enzima sono noti come **RNA polimerasi oloenzima**, questo complesso aderisce solo debolmente al DNA batterico quando i due si scontrano e un oloenzima di norma scorre rapidamente lungo la molecola di DNA fino a dissociarsi di nuovo. Quando la polimerasi scorre in una regione sulla doppia elica del DNA chiamata **promotore**, una sequenza speciale di nucleotidi che indica il punto di partenza per la sintesi di RNA, vi si lega con forza. La polimerasi, usando il fattore σ , riconosce questa sequenza di DNA prendendo contatti con le porzioni delle basi che sono esposte all'esterno dell'elica. Dopo che l'RNA polimerasi oloenzima si è legata con forza al promotore in questo modo, apre la doppia elica per esporre un breve tratto di nucleotidi su ciascun filamento. A differenza di una reazione della DNA elicasi, questa apertura limitata dell'elica non richiede l'energia di idrolisi dell'ATP. Le polimerasi del DNA subiscono invece cambiamenti strutturali reversibili che portano ad uno stato energeticamente più favorevole di quello del legame iniziale. Con il DNA svolto, uno dei due filamenti sposti di DNA agisce da stampo per formare coppie di basi complementari con i ribonucleotidi in arrivo, due dei quali sono uniti insieme dalla polimerasi per iniziare una catena di RNA. Dopo che sono stati sintetizzati i primi dieci nucleotidi circa di RNA, il nucleo dell'enzima rompe le sue interazioni con il DNA del promotore, indebolisce le sue interazioni con il fattore σ e inizia a muoversi lungo il DNA, sintetizzando RNA. L'allungamento continua fino a che l'enzima incontra un secondo segnale nel DNA, il **terminatore**, dove la polimerasi si ferma e rilascia sia lo stampo di DNA che la catena appena prodotta di RNA. Dopo il nucleo della polimerasi è stato rilasciato al terminatore, si riassocia con un fattore σ libero e forma un oloenzima che può cominciare di nuovo il processo di trascrizione.

Il processo dell'inizio della trascrizione è complesso e richiede che la RNA polimerasi oloenzima e il DNA subiscano una serie di cambiamenti di conformazione. Di questa serie di cambiamenti conformazionali, vi è un restringimento successivo dell'enzima intorno al DNA e all'RNA per assicurare che non si dissocia prima di avere finito di trascrivere un gene. Su una RNA polimerasi si dissocia prematuramente, non può riprendere la sintesi ma deve ricominciare da capo a livello del promotore. Per la maggior parte dei geni batterici un segnale di termine consiste in una fila di coppie di nucleotidi A-T preceduta da una sequenza simmetrica di DNA che quando, trascritta in RNA, si ripiega in una struttura a *forcina* mediante accoppiamento delle basi di Watson e Crick. Mentre la polimerasi trascrive attraverso un terminatore, la forcina può aiutare a "estrarre" il trascritto di RNA dal sito attivo. L'ibrido DNA-RNA nel sito attivo, che a livello dei terminatori è tenuto insieme da coppie di basi U-A, che sono meno stabili delle coppie di basi G-C perché formano due legami idrogeno per coppia di basi invece di tre, non è sufficientemente forte da tenere in posizione l'RNA e si dissocia provocando il rilascio del DNA dalla polimerasi.

I promotori batterici contengono sequenze correlate che riflettono in parte aspetti del DNA che sono riconosciuti direttamente dal fattore σ . Queste caratteristiche sono spesso riassunte sotto forma di una sequenza consenso. Una **sequenza consenso nucleotidica** viene ricavata confrontando molte sequenze con la stessa funzione di base e scegliendo il nucleotide più comune presente in ciascuna posizione. La zona promotrice, la zona che precede il primo nucleotide che deve essere trascritto dall'RNA polimerasi, identifica le *sequenze consenso*, sequenze conservate in tutti i geni. Le sequenze del DNA dei singoli promotori batterici differiscono in un modo che ne determina la forza (o numero di eventi di inizio per unità di tempo del promotore). I promotori per i geni che codificano proteine abbondanti sono molto più forti di quelli associati con geni che codificano proteine rare e le loro sequenze nucleotidiche sono responsabili di queste differenze. Come i promotori batterici, i terminatori della trascrizione comprendono anch'essi una vasta gamma di sequenze, la cui caratteristica comune più importante è quella di avere il potenziale di formare una semplice struttura a forcina di RNA. Un gene ha di norma soltanto un singolo promotore e poiché le sequenze nucleotidiche dei promotori sono simmetriche la polimerasi può sintetizzare RNA soltanto nella direzione 5'-3' e può quindi trascrivere soltanto un filamento per gene. Le sequenze genomiche rivelano che il filamento di DNA usato come stampo per la sintesi di RNA varia da gene a gene secondo la posizione nell'orientamento del promotore.

A differenza dei batteri che contengono un singolo tipo di RNA polimerasi, i nuclei eucariotici ne hanno tre: **RNA polimerasi I, RNA polimerasi II, RNA polimerasi III**. Le tre polimerasi sono strutturalmente simili fra loro e hanno in comune alcune subunità, ma trascrivono tipi diversi di geni. Le RNA polimerasi I e III trascrivono geni che codificano tRNA, rRNA e vari piccoli RNA. La RNA polimerasi II trascrive la grande maggioranza dei geni, compresi quelli che codificano proteine, e la nostra discussione si concentrerà perciò su questo enzima. Sebbene la RNA polimerasi II eucariotica abbia molte somiglianze strutturali con la RNA polimerasi batterica, ci sono parecchie differenze importanti nel modo in cui funzionano i due enzimi, due delle quali sono:

1. Mentre la RNA polimerasi batterica richiede soltanto una singola proteina addizionale (fattore σ) per iniziare la trascrizione, le RNA polimerasi eucariotiche richiedono molte altre proteine, chiamate **fattori generali di trascrizione**.
2. L'inizio della trascrizione eucariotica deve tener conto del compattamento del DNA nei nucleosomi e in forme di ordine superiore di struttura della cromatina, caratteristiche assenti nei cromosomi batterici.

I **fattori generali di trascrizione** aiutano a posizionare la RNA polimerasi correttamente sul promotore, a separare i due filamenti di DNA per permettere l'inizio della trascrizione e rilasciano l'RNA polimerasi dal promotore nella modalità di allungamento una volta che la trascrizione è cominciata. I promotori usati dalla RNA polimerasi II sono detti **TFII** (fattore di trascrizione per la polimerasi II).

I fattori di trascrizione generali eucariotici svolgono funzioni equivalenti a quelle del fattore σ nei batteri; in effetti porzioni di TFIIF hanno la stessa struttura tridimensionale delle porzioni equivalenti di σ . Il processo di assemblaggio inizia con l'attacco del fattore generale di trascrizione TFIID ad una breve sequenza di DNA a doppia elica composta soprattutto da nucleotidi T e A. Tale sequenza è nota come **TATA box** e la subunità che riconosce è detta **TBP** (TATA-Binding Protein). La TATA box è tipicamente posizionata 25 nucleotidi a monte del sito di inizio della trascrizione. Non è l'unica sequenza che segnala l'inizio della trascrizione, ma per la maggior parte dei promotori della polimerasi II è la più importante. L'attacco di TFIID provoca un'importante distorsione nel DNA della TATA box. Questa distorsione serve come punto di riferimento fisico per la posizione di un promotore attivo nel contesto di un genoma molto grande e avvicina sequenze di DNA su entrambi i lati della distorsione per permettere passaggi successivi di assemblaggio di proteine. Altri fattori vengono quindi assemblati, insieme con la RNA polimerasi II, per formare un **complesso di inizio della trascrizione** completo. Il fattore generale di trascrizione più complicato è TFIIF. Costituito da 9 subunità e segue parecchi passaggi enzimatici necessari per l'inizio della trascrizione. Dopo aver formato un complesso di inizio della trascrizione sul DNA del promotore, la RNA polimerasi II deve accedere al filamento stampo al punto di inizio della trascrizione. Il TFIIF, che contiene una DNA elicasi, rende possibile questo passaggio idrolizzando ATP e svolgendo il DNA, esponendo così il filamento stampo. Come la polimerasi batterica, la polimerasi I rimane poi nel promotore sintetizzando brevi tratti di RNA fino a che subisce una serie di cambiamenti conformazionali che le permettono di spostarsi dal promotore e di entrare nella fase di allungamento della trascrizione. Un passaggio importante di questa transizione è l'aggiunta di gruppi fosfato alla coda della RNA polimerasi, CTD o dominio C-Terminale.

Durante l'inizio della trascrizione la serina posta nella quinta posizione della sequenza ripetuta (Ser5) è fosforilata da TFIIF, che contiene una proteina chinasi in un'altra delle sue subunità. La polimerasi può allora staccarsi dal gruppo dei fattori generali di trascrizione, subendo una serie di cambiamenti conformazionali che rafforzano la sua interazione con il DNA e acquisendo nuove proteine che le permettono di trascrivere per lunghe distanze e in alcuni casi per molte ore, senza dissociarsi dal DNA. Una volta che la polimerasi II ha iniziato ad allungare il trascritto di RNA, la maggior parte dei fattori generali di trascrizione viene rilasciata dal DNA così da diventare disponibile per iniziare un altro ciclo di trascrizione con una nuova molecola di RNA polimerasi. La fosforilazione della coda della RNA polimerasi II provoca anche il caricamento sulla polimerasi di componenti del macchinario di modificazione dell'RNA e li posiziona in modo da poter modificare l'RNA mentre emerge dalla polimerasi.

Il DNA nelle cellule eucariotiche è compattato nei nucleosomi, che sono ulteriormente disposti in strutture di ordine superiore di cromatina. L'inizio della trascrizione in una cellula eucariotica richiede più proteine di quelle necessarie su DNA purificato. Proteine regolatrici dei geni note come *attivatori trascrizionali* si legano a sequenze specifiche nel DNA e aiutano ad attrarre la RNA polimerasi II al punto di inizio della trascrizione. Gli attivatori rappresentano uno dei modi principali in cui la cellula regola l'espressione dei suoi geni. L'inizio della trascrizione eucariotica richiede la presenza di un complesso proteico noto come **mediatore**, che permette alle proteine attivatrici di comunicare in modo appropriato con la polimerasi II e con i fattori generali di trascrizione. Infine richiede il reclutamento locale di enzimi che modificano la cromatina, compresi i complessi di rimodellamento della cromatina ed enzimi modificatori degli istoni. Entrambi i tipi di enzimi possono permettere maggiore accessibilità al DNA presente nella cromatina e in questo modo facilitano l'assemblaggio del macchinario di inizio della trascrizione sul DNA. Molte proteine devono assemblarsi sul punto di inizio della trascrizione per iniziare la trascrizione in una cellula eucariotica. L'ordine di assemblaggio di queste proteine non è prestabilito, ma è diverso per geni diversi. Alcuni di questi diversi complessi proteici possono interagire fra loro fuori dal DNA ed essere portati sul DNA come sottocomplessi preformati. Per iniziare a trascrivere, la RNA polimerasi II deve essere rilasciata da questo complesso di proteine, ciò richiede spesso le proteolisi in situ della proteina attivatrice.

Una volta che la trascrizione è iniziata, la RNA polimerasi si muove a scatti, fermandosi su certe sequenze e trascrivendole rapidamente altre. Le RNA polimerasi in allungamento, sia batteriche che eucariotiche, sono associate con una serie di **fattori di allungamento**, proteine che diminuiscono la probabilità che la RNA polimerasi si dissocia prima di raggiungere la fine di un gene. Questi fattori di norma si associano con la RNA polimerasi poco dopo l'inizio e aiutano le polimerasi a muoversi attraverso la grande varietà di sequenze di DNA che si trovano nei geni. Le RNA polimerasi eucariotiche devono anche lottare con la struttura della cromatina mentre si muovono lungo uno stampo di DNA e sono di solito aiutate da complessi di rimodellamento della cromatina dipendenti da ATP. Alcuni fattori di allungamento associati con la RNA polimerasi eucariotica facilitano la trascrizione attraverso i nucleosomi senza richiedere ulteriore energia. Queste proteine possono aiutare a spostare transitoriamente dimeri H2A-H2B dal nucleo dei nucleosomi, riposizionandoli quando la polimerasi si sposta attraverso il nucleosoma.

Il **superavvolgimento** del DNA rappresenta una conformazione che il DNA adotta in risposta a una tensione di superavvolgimento; viceversa, se si creano varie anse e avvolgimenti nell'elica si può creare questa tensione. In una doppia elica di DNA vi sono circa 10 coppie di nucleotidi per giro dell'elica.

La formazione di questo superavvolgimento è energeticamente favorevole perché ripristina un twist normale dell'elica nella regioni appaiate che rimangono, che altrimenti avrebbero bisogno di essere superavvolte a causa delle estremità fisse. La tensione di superavvolgimento si crea anche quando una RNA polimerasi si muove lungo un tratto di DNA che è ancorato alle sue estremità. Se la polimerasi non è libera di ruotare rapidamente, una polimerasi che si muove genera una tensione di superavvolgimento positiva nel DNA davanti ad essa e una tensione negativa dietro di essa. Per gli eucarioti si pensa che questa situazione costituisca un vantaggio: la tensione positiva di superavvolgimento davanti alla polimerasi rende più difficile aprire l'elica, ma questa tensione dovrebbe facilitare lo svolgimento del DNA nei nucleosomi, in quanto il rilascio del DNA dal nucleo di istoni aiuta a rilassare la tensione positiva di superavvolgimento. Qualunque proteina che si muova lungo un filamento di DNA di una doppia elica tende a generare una tensione di superavvolgimento. Negli eucarioti, DNA topoisomerasi rimuovono rapidamente questa tensione di superavvolgimento. Ma nei batteri una topoisomerasi specializzata chiamata DNA girasi usa l'energia di idrolisi dell'ATP per "inserire" continuamente superavvolgimenti nel DNA; mantenendo così il DNA sotto tensione costante. Questi sono superavvolgimenti negativi, avendo un senso opposto ai superavvolgimenti positivi che si formano quando una regione di elica di DNA si apre. Ogni volta che una regione di elica si apre, rimuove questi superavvolgimenti negativi dal DNA batterico, riducendo la tensione di superavvolgimento. La DNA girasi perciò rende l'apertura dell'elica del DNA nei batteri energeticamente favorevole in confronto all'apertura di un'elica di DNA che non è superavvolta. Per questa ragione in genere essa facilita quei processi genetici nei batteri, compreso l'inizio della trascrizione dalla parte della RNA polimerasi batterica, che richiedono apertura all'elica.

Gli mRNA batterici sono sintetizzati esclusivamente dalla RNA polimerasi che inizia e finisce in punti specifici del genoma. La situazione negli eucarioti è sostanzialmente diversa. La trascrizione è soltanto il primo passaggio di una serie di reazioni che comprende la modificazione covalente di entrambe le estremità dell'RNA e la rimozione di **sequenze introniche** che sono scartate dalla parte mediana del trascritto di RNA dal processo di splicing dell'RNA. Entrambe le estremità degli mRNA eucariotici sono modificate: mediante l'aggiunta di un **cappuccio** all'estremità 5' e **poliadenilazine** all'estremità 3'. Queste estremità speciali permettono alla cellula di stabilire se sono presenti entrambe le estremità di una molecola di RNA prima di esportare l'RNA dal nucleo per tradurlo in proteina. Lo splicing dell'RNA unisce le diverse porzioni di una sequenza che codifica una proteina e fornisce agli eucarioti superiori anche la capacità di sintetizzare parecchie proteine diverse dallo stesso gene. Questi passaggi di modificazione dell'RNA sono strettamente accoppiati all'allungamento della trascrizione da un meccanismo. Un passaggio chiave della transizione della RNA polimerasi II alla modalità di allungamento della sintesi di RNA è un'estesa fosforilazione della coda della RNA polimerasi II, chiamata **CTD** (dominio C-terminale).

Questo processo di fosforilazione procede gradualmente man mano che la RNA polimerasi II inizia la trascrizione, ma permette anche ad una nuova serie di proteine presenti nel punto di inizio della trascrizione, ma permette anche ad una nuova serie di proteine che funzionano nell'allungamento della trascrizione e nella modificazione del pre-mRNA di associarsi con la coda della RNA polimerasi. Alcune di queste proteine di modificazione sembrano "saltare" dalla coda della polimerasi sulla molecola nascente di RNA per iniziare a modificarla mentre emerge dalla RNA polimerasi. Così la RNA polimerasi II nella modalità di allungamento può essere vista come una fabbrica di RNA che trascrive il DNA in RNA e modifica l'RNA che produce. Completamente esteso, il CTD è lungo quasi 10 volte il resto della RNA polimerasi e, in effetti, serve da guinzaglio che trattiene una varietà di proteine nelle vicinanze finché non sono necessarie. Questo accelera la velocità delle reazioni successive.

Non appena la RNA polimerasi II ha prodotto circa 25 nucleotidi di RNA, l'estremità 5' della nuova molecola di RNA è modificata per aggiunta di un "cappuccio", che consiste in un nucleotide guaninico modificato. La reazione di aggiunta del cappuccio è svolta da tre enzimi che agiscono in successione: uno (una **fosfatasi**) rimuove un fosfato dall'estremità 5' dell'RNA nascente, un altro (una **guanil trasferasi**) aggiunge un GMP in legame inverso (5' a 5' invece di 5' a 3'); un terzo (una **metil trasferasi**) aggiunge un gruppo metilico alla guanosina. Poiché tutti e tre gli enzimi si attaccano alla coda fosforilata della RNA polimerasi a livello della serina-5 sono in posizione per modificare l'estremità 5' del trascritto nascente non appena questo emerge dalla polimerasi. Il cappuccio 5'-metile segnala l'estremità 5' degli mRNA eucariotici e questo punto di riferimento aiuta la cellula a distinguere gli mRNA dagli altri tipi di molecole di RNA presenti nella cellula. Per esempio, le RNA polimerasi I e III producono RNA privi di cappuccio durante la trascrizione, in parte perché queste polimerasi sono prive di un CTD. Nel nucleo il cappuccio si lega ad un complesso proteico chiamato CBC (Cap Binding Complex) che aiuta l'RNA ad essere modificato ed esportato correttamente. Il cappuccio 5'-metile ha anche un ruolo importante nella traduzione degli mRNA nel citosol.

Le sequenze che codificano proteine dei geni eucariotici sono interrotte da sequenze intercalate non codificanti, gli *introni*. I geni sono spezzati da tratti di sequenze codificanti, gli esoni, intervallati da sequenze intercalate o introni, così la porzione codificante di un gene eucariotico è solo una piccola frazione della lunghezza del gene. Sia gli introni che gli esoni sono trascritti in RNA. Le sequenze introniche sono rimosse dall'RNA appena sintetizzato dal processo di **splicing dell'RNA**. Solo dopo che sono avvenute le modificazioni delle estremità 5' e 3' e lo splicing questo RNA è chiamato mRNA. Ciascun evento di splicing rimuove un introne producendo attraverso due reazioni sequenziali di trasferimento di fosfato note come **transesterificazioni**, queste uniscono due esoni rimuovendo l'introne. Poiché il numero di legami fosfato ad alta energia resta lo stesso, queste reazioni potrebbero in linea di principio avvenire senza idrolisi di nucleosidi trifosfato. Il macchinario che catalizza lo splicing del pre-mRNA è complesso e consiste di 5 molecole addizionali di RNA e più di 200 proteine e idrolizza molte molecole di ATP per evento di splicing. Questa complessità è necessaria per assicurare un'elevata accuratezza dello splicing, mantenendolo allo stesso tempo flessibile per trattare gli introni presenti in una tipica cellula eucariotica. La presenza di numerosi introni nel DNA permette alla ricombinazione genetica di combinare gli esoni di geni diversi, dando modo a geni per nuove proteine di evolvere più facilmente per combinazione di parti di geni preesistenti. I trascritti di molti geni eucariotici subiscono splicing in vari modi per produrre una serie di mRNA differenti permettendo la produzione di una serie corrispondente di proteine diverse dallo stesso gene. Lo splicing permette agli eucarioti di aumentare il potenziale di codifica dei loro genomi. Lo **splicing** è una modifica del nascente pre-mRNA che avviene insieme o dopo la trascrizione nella quale gli introni vengono rimossi e gli esoni vengono uniti. Ciò è necessario per l'RNA messaggero prima che possa essere usato per produrre una corretta proteina tramite la traduzione o sintesi proteica. La maggior parte dei geni eucariotici sono costituiti da regioni codificanti, esoni, alternate a regioni non codificanti, gli introni. Gli introni devono necessariamente essere rimossi dal trascritto primario per consentire la corretta traduzione e sintetizzare la proteina corrispondente. Lo splicing dell'RNA è eseguito da molecole specializzate di RNA che riconoscono le sequenze nucleotidiche, esse specificano dove deve avvenire lo splicing e partecipano anche alla chimica dello splicing.

Queste molecole di RNA sono relativamente brevi e nella forma principale di splicing dei pre-mRNA ne sono coinvolte 5: **U1, U2, U4, U5 e U6**. Noti come **snRNA (piccoli RNA nucleari)**, sono complessati con almeno 7 subunità proteiche a formare una **snRNP (piccola ribonucleoproteina nucleare)**. Queste snRNP formano il nucleo dello **spliceosoma**, il complesso di molecole di RNA e proteine che esegue lo splicing dei pre-mRNA nella cellula. Dentro la cellula lo spliceosoma è un complesso che cattura, sottopone a splicing e rilascia l'RNA come un'unità coordinata e che subisce estesi riarrangiamenti ogni volta che produce uno splicing. Durante la reazione di splicing il riconoscimento della giunzione di splicing 5'. Del sito del punto di ramificazione e della giunzione di splicing 3' avviene in gran parte per accoppiamento di basi fra gli snRNA e le sequenze consenso di RNA nel pre-mRNA substrato. Nel corso dello splicing lo spliceosoma subisce degli spostamenti in cui una serie di interazioni di coppie di basi viene rotta e al suo posto se ne forma un'altra. Questo tipo di riarrangiamento RNA-RNA avviene parecchie volte durante la reazione di splicing e permette di controllare e ricontrollare le sequenze di RNA prima che la reazione chimica possa procedere aumentando così l'accuratezza dello splicing. L'idrolisi dell'ATP è necessaria per l'assemblaggio in più passaggi e per i riarrangiamenti dello spliceosoma. Alcune delle proteine aggiuntive che compongono lo spliceosoma usano l'energia di idrolisi dell'ATP per rompere interazione esistente RNA-RNA in modo da permettere la formazione di nuove interazioni. I riarrangiamenti RNA-RNA richiedenti ATP che avvengono nello spliceosoma si verificano all'interno delle stesse snRNP e fra le snRNP e il pre-mRNA substrato. Una delle funzioni più importanti di questi riarrangiamenti è la creazione del sito catalitico attivo dello spliceosoma. La creazione del sito catalitico dopo l'assemblaggio e il riarrangiamento dei componenti dello splicing su un pre-mRNA substrato è un modo efficace per prevenire uno splicing sbagliato.

Una volta completata la chimica dello splicing, le snRNP restano legate al cappio e il prodotto dello splicing viene rilasciato. Il disassemblaggio di queste snRNP dal cappio, richiede un'altra serie di riarrangiamenti RNA-RNA che necessitano di idrolisi di ATP, facendo così ritornare gli snRNA alla loro configurazione originale in modo da poter essere usati ancora in una nuova reazione. Al completamento dello splicing, lo spliceosoma dirige una serie di proteine a legarsi all'mRNA vicino alla posizione occupata in precedenza dall'introne. Chiamate complesso della giunzione degli esoni, queste proteine segnano il sito di un evento riuscito di splicing e, influenzano il destino successivo dell'mRNA.

Si definiscono quindi **introni** le regioni non codificanti di un gene che insieme agli esoni vengono trascritti dalle RNA polimerasi. La trascrizione di una sequenza di DNA porta quindi alla formazione di un trascritto primario, pre-mRNA, che deve essere sottoposto a splicing, il processo che porta alla rimozione degli introni alla formazione degli mRNA maturi. Gli introni sono presenti solo negli eucarioti e negli archeobatteri, ma non negli eubatteri, dove tutto l'RNA trascritto rimane nell'mRNA o nel tRNA o nel rRNA, i quali si sono diramati prima dalla linea evolutiva comune di eucarioti e archeobatteri.

Gli **esoni** costituiscono, insieme agli introni, la porzione di un gene che viene trascritto dall'RNA polimerasi durante il processo di trascrizione. Gli esoni, a differenza degli introni in seguito al processo di splicing si trovano negli mRNA maturi. Spesso, erroneamente, viene affermato che gli esoni costituiscono la parte codificante del genoma, questo è vero solo in parte poiché se e vero che tutta la sequenza che codifica una proteina deve risiedere su uno o più esoni, non è sempre vero che è un esone sia codificante, esoni non codificanti sono stati individuati in molti geni umani. L'mRNA maturo, negli eucarioti, è di solito costituito oltre che dalla parte codificante **CDS** (Coding DNA Sequence) anche da due regioni non tradotte **UTR** poste una al 5' e l'altra al 3' del trascritto. Questo comporta quindi l'esistenza di esoni completamente o parzialmente non codificanti cioè quelli che andranno a costituire le UTR del trascritto maturo.

Oltre allo splicing canonico che prevede l'eliminazione degli introni, vi è anche lo **splicing alternativo**. Lo splicing alternativo consiste nell'eliminazione degli introni e di alcuni esoni, eliminando degli esoni si producono delle proteine diverse. In questo modo avendo promotori multipli, i siti di poliadenilazione multipla o splicing alternativi è possibile avere da un unico trascritto primario, diversi mRNA e codificheranno quindi diverse proteine.