

RNA FACTORY

Il genoma di un organismo è costituito da sequenze di coppie di basi distribuite sui cromosomi. Tutte le coppie di basi del genoma vengono replicate durante la fase di sintesi del DNA, mentre soltanto alcune coppie di basi vengono trascritte in RNA. Le sequenze di coppie di basi che vengono trascritte prendono il nome di **geni**, per cui il processo di trascrizione viene anche detto **espressione genica**. All'inizio ed alla fine di ogni gene esistono sequenze dette **elementi genici regolatori**, coinvolte nella regolazione dell'espressione genica.

La trascrizione genera **quattro classi principali trascritti**, o di *molecole di RNA*:

§ RNA messaggero ↔ mRNA

§ RNA transfer ↔ tRNA

§ RNA ribosomiale ↔ rRNA

§ il piccolo RNA nucleare, *small nuclear RNA* ↔ snRNA.

Procarioti ed eucarioti posseggono entrambi l'RNA messaggero, il tRNA e l'rRNA, mentre l'snRNA si trova solo negli eucarioti. **Soltanto le molecole di mRNA vengono tradotte in proteine.**

La sintesi proteica comporta due tappe:

- **trascrizione**: è il processo di trasferimento dell'informazione dalla molecola stampo a doppia elica di DNA a una molecola di RNA a elica singola.
- **traduzione** o **sintesi proteica**: è la trasformazione in una sequenza di aminoacidi dell'informazione codificata in seno alla sequenza di basi dell'RNA messaggero.

Un gene che codifica per una molecola di mRNA, e quindi per una proteina, viene detto **gene strutturale**. Esistono trascritti di altri geni sotto forma di RNA che rappresentano prodotti finali dell'espressione genica e che non fungono come molecole messaggere tra DNA e proteina, cioè, non vengono mai tradotti in proteine. I geni che codificano per il tRNA, l'rRNA e l'snRNA appartengono a questo tipo di geni.

I geni sono quasi sempre localizzati sui cromosomi, e rimanendo nel nucleo devono essere in grado di controllare le funzioni in tutta la cellula. Essi, attraverso la trascrizione, producono dei trascritti che possono abbandonare i cromosomi ed essere coinvolti nella sintesi proteica in altre parti della cellula.

Il DNA è una struttura stabile che si ritrova pressoché immutata al termine della riproduzione cellulare, mentre **i trascritti di mRNA dovuti ai geni hanno**

un'esistenza limitata nel tempo, essendo degradati da enzimi cellulari, le *ribonucleasi*. Il mantenimento temporaneo dei trascritti di RNA è un elemento chiave nel processo di regolazione dell'espressione genica.

Nei procarioti e negli eucarioti la trascrizione avviene grazie a un processo biochimico catalizzato dalla **RNA polimerasi**. Prima che la trascrizione possa cominciare, la doppia elica di DNA deve essersi srotolata. **Soltanto una delle due eliche** di DNA viene trascritta in RNA. Questa elica è detta **elica stampo**, e l'elica che non funge da stampo, cioè quella complementare, ha la stessa polarità della catena di RNA.

Nel processo di trascrizione i **precursori dell'RNA** sono i ribonucleosidi trifosfati **ATP, GTP, CTP e UTP**, indicati come **NTP** (nucleosidi trifosfati). L'enzima che catalizza la reazione di polimerizzazione, la RNA polimerasi, specifico per la sintesi dell'RNA, utilizza soltanto NTP contenenti ribosio, e non quelli contenenti desossiribosio che è presente nei precursori del DNA. La polimerizzazione dell'RNA è molto simile a quella del DNA. Il nucleotide successivo da inserire nella catena è selezionato dalla RNA polimerasi per la sua capacità di appaiarsi alla base esposta sull'elica stampo di DNA (le catene di RNA contengono nucleotidi con la base uracile anziché timina, pertanto, quando si presenta sull'elica stampo del DNA un nucleotide A, verrà inserito sulla catena di RNA un nucleotide U).

La trascrizione del DNA è attiva durante tutti i momenti dell'interfase e corrisponde alla sintesi di una molecola di RNA a partire da un solo filamento di DNA che non sia eterocromatina, in quanto il DNA dell'eterocromatina non viene trascritto perché non contiene sequenze di geni (eterocromatina costitutiva), o perché contiene sequenze di geni inattivi (eterocromatina facoltativa). **Solo il DNA dell'eucromatina viene trascritto.**

La trascrizione **consiste dunque nel processo di lettura delle sequenze a doppia elica del DNA in sequenze a singola elica di RNA**: mRNA, tRNA, e rRNA in procarioti ed eucarioti, e nei soli eucarioti si verifica anche la sintesi di snRNA.

La trascrizione è un processo che presenta caratteristiche simili in procarioti ed eucarioti: il DNA si denatura e la RNA polimerasi catalizza la sintesi di una molecola di RNA in direzione $5' \rightarrow 3'$. **Sequenze promotrici e terminatrici** stabiliscono rispettivamente dove debba cominciare e finire la trascrizione.

In *Escherichia coli* la sola RNA polimerasi trascrive tutte le diverse classi di RNA della cellula e cioè l'mRNA, il tRNA e l'rRNA. Nei nuclei delle cellule eucariotiche esistono **tre RNA polimerasi** con funzioni diverse: la *RNA polimerasi I* trascrive tre dei quattro rRNA, la *RNA polimerasi II* trascrive gli mRNA e alcuni snRNA mentre la *RNA polimerasi III* trascrive il quarto rRNA, i tRNA e i rimanenti snRNA.

Tutte e tre sono enzimi di notevoli dimensioni (>500 kD, 12 subunità). Sono in grado di effettuare sintesi di RNA dipendente da stampo, ma non sono in grado di iniziare

in maniera selettiva al promotore. Alcune sub unità sono comuni a tutte e tre le RNA polimerasi.

L'RNA Pol.II è formata da 2 subunità grandi (Large) L (128-150 kD) e L' (160-220 kD) e da 12-15 subunità più piccole. La più grande sub unità in Pol.II, L', ha un dominio carbossi-terminale (CTD) che è formato da ripetizioni multiple di una sequenza consenso di 7 aminoacidi (Tyr-Ser-Pro-Thr-Ser-Pro-Ser).

In tutti gli eucarioti l'attività di RNA Pol.II è inibita da basse concentrazioni di α -amanitina, un octapeptide biciclico (L' α -amanitina è un componente tossico del fungo velenoso *Amanita phalloides*, blocca la sintesi di RNA da parte della RNA polimerasi II, e, a concentrazioni più alte, della RNA polimerasi III; non ha invece alcun effetto sulla trascrizione nei batteri).

Il controllo della Trascrizione della RNA Polimerasi II

La regolazione della trascrizione dei geni eucariotici codificanti proteine è un processo orchestrato che richiede il funzionamento concertato di molte proteine o di fattori di trascrizione.

Questi fattori possono essere classificati in tre gruppi:

(1.) regolatori trascrizionali leganti sequenze specifiche di DNA (ad es. **attivatori e repressori**) i quali si legano ad elementi prossimali ai promotori e/o sequenze regolative distali (ad es. **enhancer e silenziatori**) e modulano il livello di trascrizione di specifici geni bersaglio in maniera tessuto- e sviluppo-specifica od in risposta a stimoli fisiologici od ambientali;

(2.) fattori di trascrizione generali/basali che sono ubiquitari ed includono l'RNA Polimerasi II (Pol II) ed una serie accessoria di **fattori generali iniziatori della trascrizione (GTFs)** che si legano al "core" degli elementi promotori del DNA (come la TATA box e gli iniziatori) e permettono il reclutamento specifico di Pol II sul "core" del promotore dei geni di classe II;

(3.) cofattori/co-regolatori della trascrizione (**coattivatori e corepressori**), i quali interagiscono con i regolatori e giocano un ruolo essenziale nella mediazione e nella facilitazione dell'effetto dei regolatori sul macchinario della trascrizione generale/basale sia attraverso interazioni fisiche con i GTFs e/o Pol II o indirettamente attraverso modificazioni della struttura cromatinica.

Le interazioni cooperative proteina-proteina e proteina-DNA rivestono un ruolo essenziale nel legame specifico degli attivatori sugli aumentatori (enhancers) – dando come risultato la formazione di complessi nucleoproteici specifici chiamati corpi aumentatori (enhanceosomi) (Merika e Thanos, 2001) – e nel preciso assemblaggio

dei GTFs e Pol II in un complesso di pre-inizio stabile su “core” del promotore (Orphanides *et al.*,1996). In contrasto ai regolatori sequenza-specifici che generalmente legano il DNA come singola proteina o come complesso dimerico, vari componenti del macchinario generale/basale della trascrizione esistono come complessi multiproteici stabili in soluzione e si possono inoltre pre-assemblare al di fuori del DNA in un grande oloenzima con molte subunità. Questo fenomeno è ancora più evidente per molti gruppi più recentemente caratterizzati di cofattori e co-regolatori della trascrizione.

L’Inizio della trascrizione

I fattori di trascrizione generali (GTF), sono necessari per l’inizio della sintesi dell’RNA di tutti i geni di classe II. Essi formano con la RNA Pol. II un complesso in prossimità del punto di inizio della trascrizione, determinando il sito d’inizio; questo complesso è chiamato **complesso basale di trascrizione**.

I più studiati sono i seguenti:

- TBP (l’unico che lega il DNA in modo sito-specifico; con i fattori associati a TBP – TAFs forma TFIID)
- TFIIA
- TFIIB
- TFIIF
- TFIIE
- TFIIH

I GTFs sono quasi tutti proteine multimeriche, sono altamente conservati dal lievito ai mammiferi. Il loro assemblaggio su un “core promoter” avviene con una sequenza precisa, almeno in vitro.

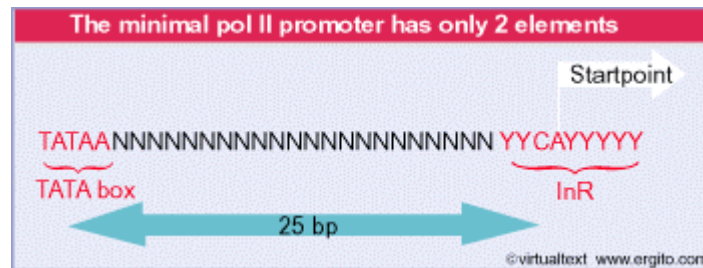
L’assemblaggio è stato studiato in vitro tramite esperimenti di trascrizione in vitro con componenti purificate e tramite footprint con DNasi I, saggio del cambio di mobilità elettroforetica (Electrophoretic Mobility Shift Assay, EMSA), in vivo tramite il saggio di immunoprecipitazione della cromatina (Chromatin Immunoprecipitation assay, ChIP).

Ad eccezione di TFIIB, che è un singolo polipeptide di 35 kDa, gli altri GTFs (isolate in cellule di mammifero) e Pol II risultano essere formati da complessi multiproteici stabili che comprendono da 2 a 14 differenti proteine. La RNA Pol II stessa è un grande enzima a più subunità che è stato caratterizzato in lievito, ed è composto da 12 differenti proteine delle quali 5 (ossia RPB5, RPB6, RPB8, RPB10, ed RPB12) sono condivise con le altre due RNA polimerasi eucariotiche, Pol I e Pol III.

La struttura cristallografica ad alta risoluzione della Pol II di lievito a 10 subunità, libera ed in fase di trascrizione, è stata recentemente descritta (Cramer *et al.*, 2001; Gnatt *et al.*, 2001).

Malgrado la sua complessità, Pol II da sola non è capace di riconoscere i promotori e di iniziare accuratamente la trascrizione.

L'inizio specifico della trascrizione da parte di Pol II richiede il riconoscimento da parte delle GTFs di vari elementi nel promotore – che includono la TATA box, l'inziatore (INR), l'elemento di risposta al TFIIB (BRE) e gli elementi DPE (Downstream Promoters Elements) – e l'assemblaggio cooperativo dei GTFs e Pol II in un **complesso di pre-inizio (PIC)** sul “core” del promotore (Orphanides *et al.*,1996).



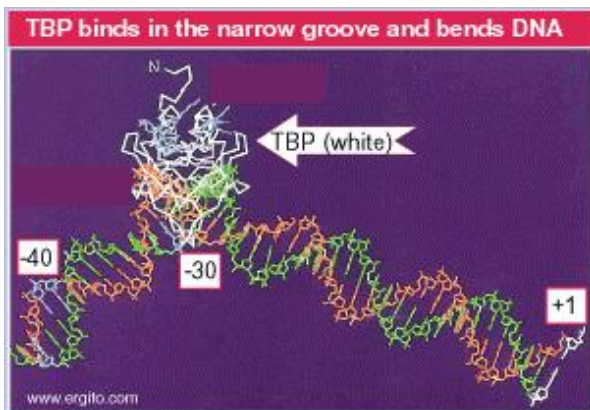
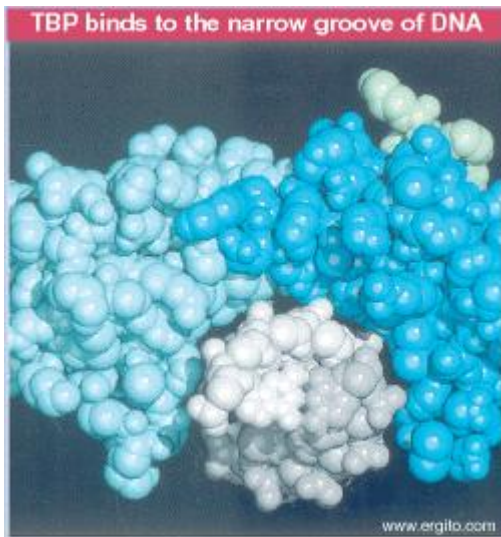
Il “core” del promotore viene definito come la minima sequenza promotrice di DNA che è necessaria e sufficiente a dirigere l'inizio specifico della trascrizione ad parte di Pol II e dei GTFs *in vivo*.

Il primo passo coinvolto nella formazione del PIC e dell'inizio della trascrizione nel “core” del promotore, contenente la TATA box e l'inziatore (INR), è **il legame di TFIID al “core”**.

Il TFIID umano è formato da due classi di polipeptidi:

- la **TBP (TATA binding protein)** è una piccola proteina (30kD) che lega la TATA box; è un monomero che si ripiega a “sella di cavallo” le cui due metà contattano il DNA lungo il solco minore causando una notevole distorsione (bending).
- i TAF (TBP associated factors), con pesi molecolari tra i 30 e i 250 kD. I TAF sono considerati co-attivatori (cioè non sono necessari per la trascrizione basale *in vitro*).

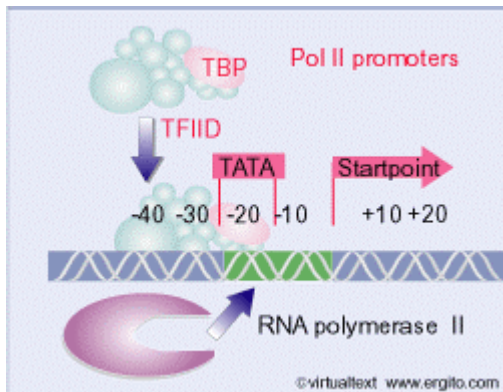
TFIID raggiunge le dimensioni di 800 kD, essendo formato da TBP e da circa 11 TAF.



La TBP è presente in tutti i fattori di posizionamento, nei promotori della RNA Pol.III TFIIB (che contiene la TBP) si lega a TFIIC; nei promotori della RNA Pol.I il fattore di legame al “core” (che contiene la TBP) è facilitato nel legame al DNA dalla presenza di UBF1.

Nei promotori della RNA Pol. II TBP lega direttamente la sequenza TATA – per gli altri promotori il legame al DNA è compito di altre proteine.

Compito della TBP è, in maniera diretta o indiretta, il posizionamento della RNA Pol.(I, II o III) sul sito di inizio.



La TBP lega il DNA in maniera atipica: nel solco secondario (minore), mentre tipicamente le proteine che legano il DNA contattano la doppia elica nel solco principale (maggiore).

La TBP contatta la doppia elica da un lato, la superficie interna della TBP lega la doppia elica, mentre la più ampia superficie esterna è disponibile per l'interazione con altre proteine.

TBP si lega nel solco secondario e piega il DNA di 80°; la TATA box si piega verso il solco maggiore, allargando il solco minore, mentre alcuni dei TAF interagiscono con l'INR (al sito di start della trascrizione/+1) ed a sequenze di DNA aggiuntive situate a valle (a circa +35), che in certi geni includono gli elementi DPE (Wu *et al.*, 2001).

La curvatura del DNA permette agli altri fattori di trascrizione ed alla RNA polimerasi di contattare il DNA in maniera più intima di quanto non sarebbe possibile con una molecola lineare.

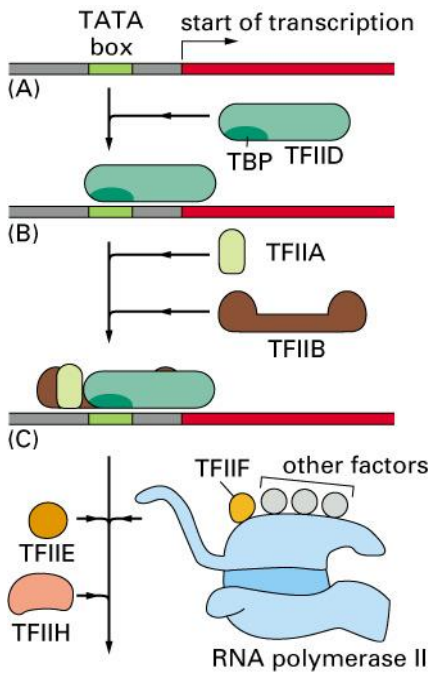


Figure 6-16 part 1 of 2. Molecular Biology of the Cell, 4th Edition.

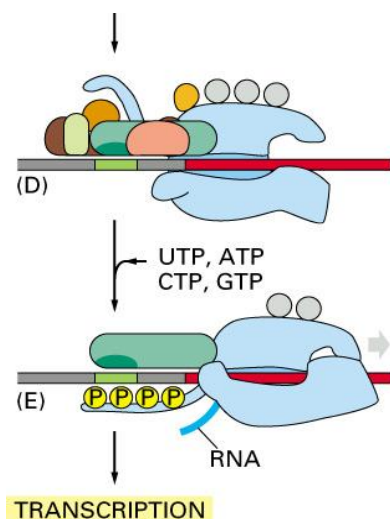
L'assemblamento del complesso di inizio procede a stadi seguendo una determinata sequenza.

- Il primo evento è il legame di TFIID alla TATA.
- Il legame di TFIIA aumenta la regione di DNA contattata da TFIID.
- Poi si lega **TFIIB**. TFIIB contatta il solco secondario a valle della tata box ed il solco principale a monte della TATA. TFIIB probabilmente fornisce la superficie che sarà contattata dalla polimerasi, posizionandola correttamente.

Il legame di TFIID al “core” del promotore è **stabilizzato da TFIIA e TFIIB** in differenti modi. TFIIB stabilizza il complesso TBP-TATA attraverso contatti sia con la TBP che con sequenze di DNA (come BRE) fiancheggianti l'elemento TATA piegato (Tsai e Silger, 2000). Anche TFIIA stabilizza le interazioni TBP-TATA contattando la TBP ed il DNA a monte dell'elemento TATA (Geiger *et al.*, 1996).

Tuttavia, in contrasto rispetto a TFIIB, TFIIA risulta superfluo per la trascrizione basale diretta in sistemi purificati *in vitro*. Comunque **TFIIA ha un effetto stimolatore** sulla trascrizione basale TFIID-dipendente nei “core” dei promotori contenenti TATA ed INR che richiedono i TAF per una ottimale attività basale e per il sinergismo fra TATA ed INR (Hansen e Tjian, 1995).

Inoltre TFIIA è essenziale per la trascrizione dei “core” di promotori mancanti della TATA box ma contenenti INR che richiedono anche TFIID/TAF. In più, in sistemi di trascrizione meno purificati, TFIIA ha una funzione di anti-repressione competendo con i cofattori che interagiscono negativamente con la TBP (NCs), come NC2/Dr1-DRAP1, che prevengono l'associazione di TFIIB con il complesso TATA-TBP. In questo modo, la necessità della presenza di TFIIA nella trascrizione basale è specifica riguardo il “core” del promotore e sembra collegata all'associazione dei TAF (o NCs) con la TBP e alla funzione promotore-selettiva del complesso TFIID. In accordo con questo, TFIIA ha mostrato di indurre/riorganizzare o stabilizzare le interazioni dei TAF con l'INR e le sequenze a valle (Emami e Smale, 1997). Oltre le loro attività selettive per i “core” dei promotori, TFIIA e i TAF fungono anche da coattivatori.



La formazione di un complesso **TFIID-IIA-IIB-DNA** (o in maniera minimale **TBO-IIB-DNA**), permette il reclutamento susseguente della **Pol II**.

L'incorporazione **addizionale di TFIIE e di TFIIF** è richiesta per l'apertura (melting) del DNA dovuta alla formazione del complesso, e l'efficiente inizio di trascrizione di un DNA lineare *in vitro*.

TFIIE ricopre un ruolo nel reclutamento di TFIIH e nello stimolo delle sue attività. **TFIIH** è un complesso di nove subunità con una architettura ad anello (Schultz *et al.*, 2000).

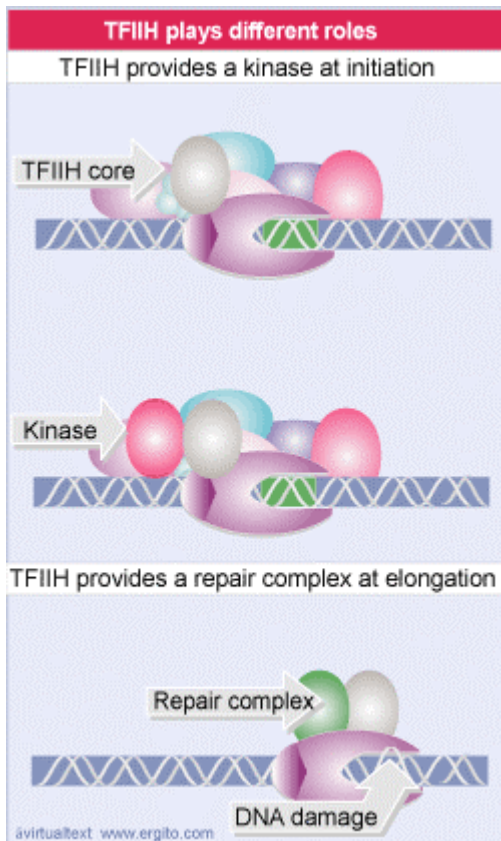
TFIIH ha varie attività enzimatiche:

- ✓ ATPasi (fonte di energia)
- ✓ Elicasi (per separare i due filamenti di DNA)
- ✓ Chinasi (fosforila il **CTD**, dominio carbossi terminale, della più grande subunità della Pol II, un evento che è ritenuto faciliti la liberazione del promotore da parte dei componenti del PIC).

TFIIE e TFIIH, usando idrolisi di ATP, aprono i due filamenti di DNA e permettono alla RNA Pol.II di iniziare la trascrizione.

La polimerasi sintetizza il primo legame fosfodiesterico tra i primi due nucleotidi. A questo punto l'attività chinasi di TFIIH fosforila il CTD, e questo permette alla polimerasi di iniziare la fase di elongazione.

Nell'allungamento la Pol.II si libera di molti TF e recluta **fattori di allungamento (TFIIS e Hspt5)**. TFIIS stimola l'allungamento e facilita la correzione.



TFIID , oltre che nell'inizio della trascrizione, è coinvolto anche nella riparazione del DNA.

Il fattore esiste in forme diverse a seconda della funzione che deve svolgere. La forma base di TFIID è un "core" composto da 5 subunità. Nella forma attiva nella trascrizione, il "core" è associato ad un complesso che ha attività chinasi (fosforila il CTD della polimerasi).

In una forma alternativa, il "core" è associato ad un gruppo di proteine coinvolte nella riparazione:

- ✓ XPC riconosce il DNA danneggiato, quando la polimerasi si ferma a causa di un danno
- ✓ XPG e XPF: endonucleasi.

Il complesso chinasi e quello della riparazione possono associarsi e dissociarsi reversibilmente, TFIID si dissocia dopo 50 bp di trascrizione, e si riassocia a siti danneggiati.

I geni codificanti le subunità di TFIID (XPB e XPD) sono mutati in tre malattie genetiche umane: lo xeroderma pigmentoso, la sindrome di Cockayne e la trichotiodistrofia (Hoeijmakers *et al.*, 1996; de Laat *et al.*, 1999).

Il Complesso di Pre-Inizio (PIC) completo è dunque un grande complesso nucleoproteico composto da 43 proteine distinte con una massa totale di più di 2,2 MDa (che non include le circa 24 subunità del complesso mediatore).

Dopo la formazione del complesso d'apertura, l'inizio della trascrizione e lo sgombrò del promotore, La Pol II con la CTD fosforilata (Pol II₀) in associazione con **TFIIF allunga il gene a valle**, mentre TFIIA e TFIID rimangono legati al core del promotore e TFIIB, TFIIE e TFIID vengono rilasciati (Zawel *et al.*, 1995).

Uno studio recente (nel quale estratti nucleari di lievito sono utilizzati come risorsa per i fattori di trascrizione) ha mostrato che, dopo l'inizio della trascrizione e la "fuga" di Pol II, i componenti del Mediatore possono rimanere associati con il "core" del promotore assieme con TFIIA, TFIID, TFIIE e TFIID formando una impalcatura che è stabilizzata dagli attivatori. Questo suggerisce che un secondo "turno" di inizio di trascrizione non debba richiedere un reclutamento *de novo* di TFIIE, TFIID o

dell'oloenzima Mediatore/Pol II, ma debba soltanto richiedere la re-incorporazione di TFIIB, TFIIF e Pol II (Yudkovsky *et al.*, 2000).

La Pol II che si trova a re-iniziare potrebbe essere una forma defosforilata Pol IIA che viene generata durante o dopo la terminazione della trascrizione tramite una fosforilasi CTD-specifica (modello generalmente accettato) o, alternativamente, una forma Pol IIO può essere capace di re-iniziare la trascrizione al “core” del promotore in presenza di attivatori che si legano a promotori a monte (Cassaño *et al.*, 2000).

La sequenza di eventi che portano alla formazione del PIC descritta prima è governata dal legame iniziale di TBP/TFIID alla TATA box. Tuttavia, molti geni non hanno un elemento TATA ed i “core” dei promotori mancanti di TATA e contenenti INR possono dirigere un inizio accurato della trascrizione tramite la Pol II *in vivo*.

I fattori ed i meccanismi coinvolti nella formazione del PIC sono tuttora poco caratterizzati. I promotori mancanti dell'elemento TATA richiedono anche TFIID (Zhou *et al.*, 1992). Comunque, in contrasto con i promotori contenti la TATA box che possono dirigere la trascrizione indipendente da attivatori (basale) in presenza della TBP, i promotori mancanti di TATA in cellule di mammifero richiedono TAF all'interno del complesso TFIID e TFIIA per la trascrizione basale (Roeder, 1998).

Inoltre, una mutazione sulla superficie di TBP che lega il DNA che inibisce la trascrizione da promotori di mammifero contenti la TATA box *in vivo* ed *in vitro*, non influenzano la trascrizione basale da “core” di promotori mancanti di TATA ma contenenti INR (Martinez *et al.*, 1995).

Questi esperimenti dimostrano l'esistenza di percorsi alternativi per il reclutamento funzionale di TFIID che è selettivo per i “core” dei promotori. Si ritiene interessante il fatto che TAF250 e TAF150 umani, in associazione, si legano debolmente ma direttamente all'INR in maniera sequenza specifica (Chalkley e Verrijzer, 1999) ed inoltre che le TAF di *Drosophila* con struttura simile all'istone H3/H4 siano state legate covalentemente (crosslinked) alla sequenza DPE (elemento a valle del promotore) presente dopo l'INR (a +30) in dei geni contenenti o meno la TATA box (Burke e Kadonaga, 1997). Così, i contatti TAF-DNA possono costituire un'alternativa all'interazione TBP-DNA e sono sufficienti per il reclutamento di TFIID in certi geni mancanti della TATA box ma contenenti INR che contengono anche una DPE.

È ora chiaro che differenti passaggi limitanti esistono nell'assemblaggio del PIC e nell'inizio della trascrizione nei differenti “core” dei promotori, e che sebbene i GFTs siano ubiquitari e generalmente richiesti per la trascrizione di molti geni di classe II, hanno anche funzioni selettive per i “core” dei promotori o di interazione con cofattori (come NCs, TAF, TICs) che possono modulare la loro attività in maniera selettiva.

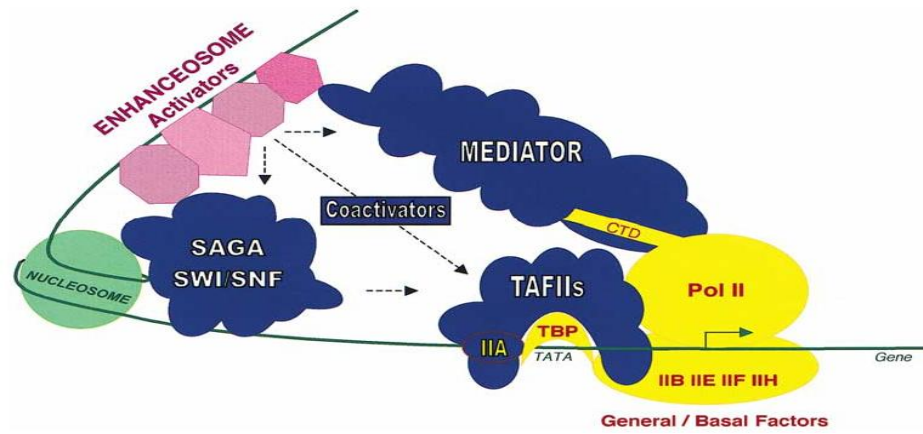


Figura 6. L'induzione della trascrizione genica tramite degli attivatori coinvolge vari coattivatori ed interazioni proteina-proteina. I coattivatori (blu) sono reclutati dagli attivatori legati ai promotori (poligoni rosa) per rimodellare la struttura cromatinica (nucleosoma verde) e/o stimolare il reclutamento e l'attività del macchinario basale della trascrizione (giallo).

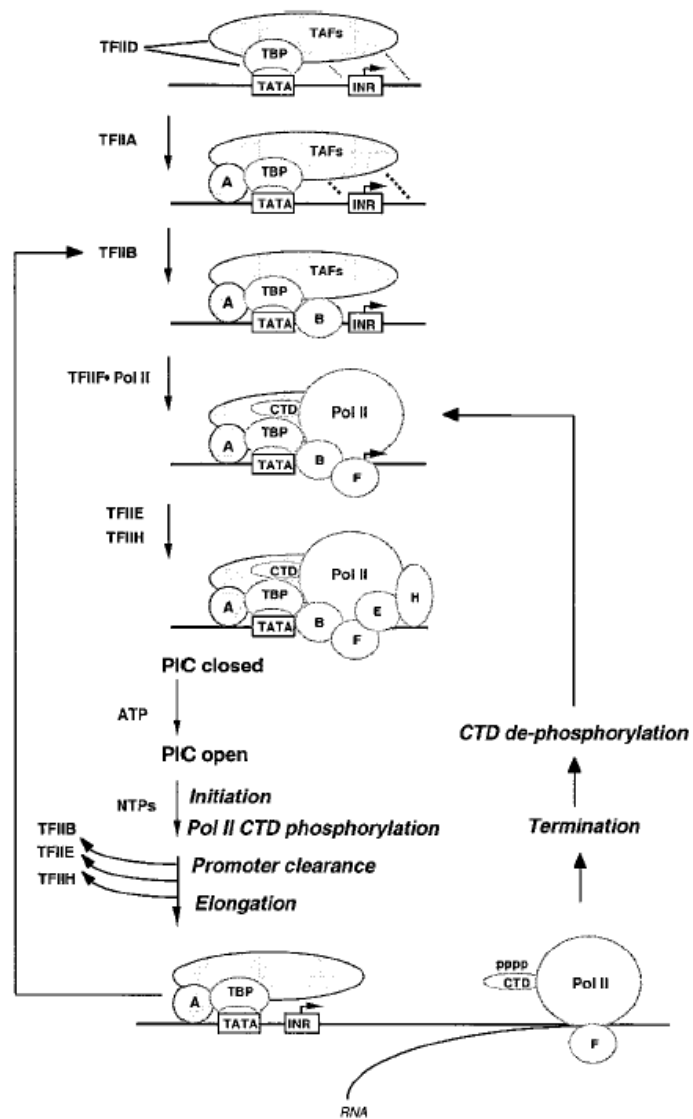


Figura 7. Modello dell'assemblaggio ed il funzionamento del complesso di pre-inizio (PIC).

Bibliografia:

<http://atlasgeneticsoncology.org/Educ/TFactorsID30020IS.html>

<http://www.unicz.it/didattica/lauree/Biotecnologie/dispense/biologiamolecolare>

<http://www.summagallicana.it/Volume2/B.III.06.2.htm>

<http://med.unibs.it/chiamica/biol%20mol/lezione7.ppt>

<http://ncbi.nlm.nih.gov/books/bv.fcgi?rid=mboc4.figgrp.1010>

http://www.bioinformatica.unito.it/downloads/dispense_biol-mol/lezione4.pdf

http://hal9000.cisi.unito.it/wf/DIPARTIMEN/Genetica, /Attivit--D1/Biologia-m/4- Trascrizione-1_06-07.ppt_cvt_file/frame.htm

<http://users.unimi.it/poletti/biologia/TESTI/Trascrizione/trascrizione18.html>

http://www.pesolelab.it/sections/05_Courses/04_AA2007-2008/03_BioMol/02_MatDid/L6.pdf

http://www.bioinformatica.unito.it/downloads/dispense_biol-mol/lezione4.pdf