

## ARCHITETTURA MOLECOLARE DI FATTORI DI TRASCRIZIONE DEL DNA

MARCO NARDINI (\*)

Nota presentata dal m.e. Martino Bolognesi  
(Adunanza del 26 novembre 2020)

SUNTO. – I fattori trascrizionali sono proteine coinvolte in uno dei processi molecolari più importanti nelle cellule viventi, che è l'espressione dei geni attraverso un meccanismo chiamato "trascrizione". I fattori trascrizionali, la loro struttura e la loro funzione sono l'argomento del seminario che include anche una panoramica su quale è il livello di conoscenza attuale, quali sono le potenzialità e le problematiche in questo campo di ricerca e le possibili applicazioni.

\*\*\*

ABSTRACT. – Transcription factors are proteins involved in one of the most important molecular processes in living cells: the expression of genes through a mechanism called "transcription". Transcription factors, their structure, and function are the topic of the seminar which also includes an overview of the current level of knowledge, the potentialities and problems in this research field, and the possible applications.

I fattori trascrizionali sono proteine che giocano un ruolo fondamentale nel processo dell'espressione dei geni attraverso il meccanismo della "trascrizione". Per capire cosa si intende per "espressione di un gene" bisogna prima spiegare quello che nel campo viene definito come il "dogma centrale della biologia molecolare" che spiega il meccanismo di sintesi delle proteine nelle cellule. L'informazione genetica necessaria

---

(\*) Laboratorio di Biologia Strutturale, Dipartimento di Bioscienze, Università degli Studi di Milano, Italy. E-mail: marco.nardini@unimi.it

per sintetizzare le proteine è contenuta nel DNA (l'acido desossiribonucleico a doppia elica), in particolare in parti del DNA chiamati "geni", ma questa informazione non viene trasferita direttamente dal DNA alle proteine, bensì attraverso un passaggio intermedio. Il "dogma centrale della biologia molecolare" afferma che il DNA dirige la propria trascrizione in una sequenza complementare di un'altra molecola nucleotidica, questa volta a singolo filamento, chiamata RNA (acido ribonucleico), che viene poi tradotta in una sequenza di amminoacidi che va a formare una proteina. Quindi, l'espressione di un gene essenzialmente è un meccanismo di trasferimento di informazione biologica che avviene all'atto della trascrizione del DNA in RNA. Questo meccanismo è molto complesso ed estremamente importante per molti processi fondamentali in biologia. In altre parole, cosa fanno le cellule, come funziona un tessuto e come un organismo sopravvive e si riproduce dipende dall'espressione dei geni ed il primo passo di questo processo è la trascrizione. La regolazione dell'utilizzo dell'informazione genetica che è codificata dal DNA delle cellule viene effettuata attraverso dei "fattori" proteici detti "fattori di trascrizione", coinvolti per esempio nello sviluppo embrionale, nel differenziamento cellulare e nel determinare il destino delle cellule.

Capire i meccanismi molecolari che controllano la trascrizione nell'uomo o in qualunque altro organismo può consentire di razionalizzare cosa accade nel caso di malattie (tumori, diabete, Parkinson, ecc.), con la speranza di poter sviluppare ed applicare per queste complesse patologie dei trattamenti farmaceutici più specifici e anche di sviluppare degli approcci diagnostici più accurati e rapidi.

In questo campo di ricerca, uno dei più importanti risultati scientifici ottenuti negli ultimi 20 anni è stato il sequenziamento del genoma umano, che contiene circa 3 miliardi di coppie di basi di DNA, che corrispondono a circa 22000 geni. Un risultato molto importante, ed in parte inatteso, è che in realtà il numero di geni presenti nel genoma umano non è molto diverso da quello presente in genomi di altri organismi, anche molto più semplici. Parte della spiegazione a questo apparente paradosso è che solo circa il 3% del DNA dell'uomo è formato da geni che saranno poi trascritti e tradotti in proteine, mentre il rimanente 97% del DNA non codifica per proteine. Quindi, di fatto, la maggior parte del DNA nei nostri cromosomi si dice essere "non codificante". Tuttavia, queste regioni del DNA sono estremamente importanti perché controllano se, e come, un gene viene attivato (cioè trascritto)

oppure no. In particolare, le regioni non codificanti contengono dei piccoli frammenti di DNA, costituiti da poche coppie di basi, che sono chiamate sequenze regolatorie che sono riconosciute e legate dai fattori trascrizionali, detti “sequenze specifiche”, che controllano se un gene è espresso oppure no in particolari condizioni. Quindi la differenza fra una cellula ed un'altra, fra un organismo ed un altro o la risposta che una certa cellula può dare ad un certo stimolo non è tanto legata al numero dei geni presenti, ma al modo in cui la loro trascrizione viene coordinata e regolata dai fattori trascrizionali.

Per capire come avviene l'azione regolatoria dei fattori di trascrizione è necessario ricordare che negli eucarioti la trascrizione del DNA in RNA avviene grazie all'enzima RNA polimerasi II (RNAPol-II). Tale enzima è molto complesso (550 kDa, 12 subunità) e costituisce una macchina molecolare in grado di separare i due filamenti del DNA, di leggerne uno, detto DNA stampo o antisense, e di sintetizzare una molecola di RNA a singolo filamento che possiede le basi complementari al filamento di DNA stampo. Dopodiché, mentre compie questa operazione di copiatura scorrendo lungo il gene, ricomponde la doppia elica di DNA copiata dietro di sé finché non giunge alla fine del gene. Tuttavia, la RNAPol-II ha bisogno di aiuto per portare a termine il suo compito. In particolare da solo questo enzima non è in grado di distinguere le regioni del DNA codificanti, cioè quelle dove sono localizzati i geni, rispetto a quelle non codificanti. Questo significa che devono esistere degli altri fattori in grado di dirigere la RNAPol-II nel posto giusto del DNA ed al momento giusto per iniziare la trascrizione dei geni giusti, in modo che ciascuna cellula possa funzionare in modo corretto. Questi fattori, sono proprio i fattori di trascrizione che riconoscono delle brevi sequenze di DNA presenti nelle zone non codificanti, le legano in modo specifico e “accendono” o “spengono” la trascrizione dei geni attraverso la RNAPol-II.

Da un punto di vista meccanicistico, quello che accade affinché un gene possa essere trascritto, è la formazione di un complesso di pre-inizio (detto PIC) formato da subunità proteiche multiple che si posiziona sul promotore del gene, cioè sulla parte del DNA che precede il gene che deve essere espresso, e che recluta la RNAPol-II. Nei promotori di molti geni, anche se non tutti, è presente una sequenza di DNA chiamata TATA-box, che rappresenta il sito di legame di una proteina particolare chiamata TATA-binding protein (TBP), contenuta nel complesso trascrizionale TFIID. Il legame di TFIID al promotore è il primo

passo che innesca il legame di altri fattori trascrizionali detti “generalisti” che contribuiscono a posizionare correttamente la RNAPol-II. La TBP è una proteina interessante da un punto di vista strutturale perché ha una caratteristica forma a sella, perfettamente complementare alla forma della doppia elica di DNA che lega sul TATA-box.

Come già anticipato, l’attivazione della trascrizione è regolata dall’opportuno posizionamento di altri fattori trascrizionali, sequenza specifici, su regioni regolatorie del DNA localizzate non solo sul promotore del gene, cioè vicini al gene da trascrivere, ma anche in regioni distanti (enhancer prossimali) o molto distanti (enhancer distali). Questo aspetto è interessante perché indica che, affinché questi fattori trascrizionali possano influenzare in maniera diretta o indiretta la trascrizione, sia come attivatori che come repressori, è necessario che il DNA faccia dei loop avvicinando così al promotore le zone di DNA regolatorie con i fattori trascrizionali legati.

La grande complessità strutturale e regolatoria della trascrizione negli eucarioti dipende anche dal fatto che la macchina trascrizionale deve muoversi in un contesto in cui il DNA non è in una forma estesa e sempre facilmente accessibile. Infatti nelle cellule eucariote il materiale genetico è organizzato in una struttura complessa e compatta composta da DNA e proteine chiamata cromatina. L’unità fondamentale della cromatina è il nucleosoma, in cui la doppia elica di DNA si arrotola attorno a proteine chiamate istoni. La conseguenza di tutto ciò è che è possibile confinare tutto il DNA in un piccolo nucleo. Per dare un’idea, se si estendesse tutto il DNA contenuto nel nucleo di una cellula, esso sarebbe lungo più di un metro. D’altra parte però, una forma così condensata rende complicato avere accesso all’informazione genetica contenuta nel DNA da parte della macchina trascrizionale. Ci devono essere, quindi, anche delle proteine in grado di modificare gli istoni e rimodellare la cromatina per rendere i geni accessibili alla macchina trascrizionale.

Da un punto di vista strutturale, i fattori trascrizionali e, più in generale, le proteine che legano il DNA e influenzano la trascrizione, hanno varie forme e dimensioni ma sono caratterizzate da alcuni tratti comuni. Nella maggior parte dei casi tali proteine sono modulari, cioè contengono regioni a struttura ben definita ed isolabile detti domini. Nella forma più semplice, i fattori trascrizionali sono costituiti da due domini, uno che riconosce una sequenza di DNA e la lega, l’altro che attiva o reprime la trascrizione, cioè è dotato di attività regolatoria e può essere utilizzato, per esempio, per far comunicare fra loro fattori

trascrizionali diversi. Di solito nel dominio che lega il DNA ci sono delle regioni strutturali che tendono ad essere complementari con la struttura del DNA in termini sia di forma sia di carica. Per esempio, contengono regioni caratterizzate da un alto numero di amminoacidi carichi positivamente che si interfacciano con la carica negativa del DNA ottimizzando così l'interazione elettrostatica attrattiva fra le due macromolecole. Ci sono inoltre regioni del fattore trascrizionale che si inseriscono nelle spire dell'elica del DNA, di solito nel solco maggiore, e riconoscono le basi con cui interagire, cioè capiscono dove posizionarsi sul DNA "leggendo" la sua sequenza nucleotidica. Di solito queste strutture possono essere piccoli domini che contengono ioni, come gli zinc finger che contengono zinco, o presentano delle eliche, più o meno lunghe. In alcuni casi la struttura dei fattori trascrizionali è più complessa, come nel caso della proteina NF-Y, un fattore trascrizionale che lega il motivo CCAAT-box sul DNA e che è formata da tre subunità. Due di queste subunità hanno la stessa struttura presente negli istoni del nucleosoma e modificano la struttura del DNA allo stesso modo, piegandolo di ben 80° circa. Quindi NF-Y rappresenta un esempio in cui il legame del fattore trascrizionale altera anche fisicamente la struttura del DNA, lo piega e allarga il solco minore del DNA (e non quello maggiore come accade di solito nei fattori trascrizionali) in cui poi si inserisce la terza subunità che legge la sequenza specifica del CCAAT-box. Questa terza subunità, che interagisce con le basi nucleotidiche del CCAAT-box, ha la forma ad elica. Avere un fattore trascrizionale che si posiziona nel solco minore del DNA è importante perché può permettere al DNA di legare un altro fattore trascrizionale nel solco maggiore adiacente e quindi avere un livello di cooperatività positiva o negativa fra i due fattori trascrizionali, fornendo un ulteriore grado di complessità nella regolazione della trascrizione.

Lo studio strutturale dei complessi macromolecolari proteine/DNA, che sono in generale molto grandi ma anche molto dinamici richiede l'utilizzo di tecniche sperimentali di biologia strutturale diversificate e complementari, come per esempio la cristallografia a raggi X, che consente di raggiungere alta risoluzione su campioni cristallini, ma anche la più recente crio-microscopia elettronica a singola particella che permette di visualizzare complessi più grandi utilizzando campioni in soluzione. Nel 2017 il premio Nobel per la Chimica è stato assegnato a Jacques Dubochet, Joachim Frank e Richard Henderson proprio per lo sviluppo e l'applicazione di questa tecnica su campioni

biologici. E sempre nel 2017 nel dipartimento di Bioscienze dell'Università di Milano (in collaborazione con la Fondazione Invernizzi) è stato installato e reso operativo il primo crio-microscopio elettronico in Italia. A distanza di alcuni anni molti altri microscopi di questo tipo si stanno installando un po' in tutta Italia, proprio per il loro grande potenziale. Questi microscopi consentono di vedere questi complessi macromolecolari e capire come funzionano a livello atomico, anche nella loro dinamicità, visualizzando con eccezionale nitidezza oggetti della dimensione di un centesimo di micron, come per esempio il nucleosoma. Questa metodologia apre una nuova era in questo campo di ricerca così complesso.