

# Metodi spettrali per l'analisi dei promotori del DNA

Lucia Pettinato [plucix@yahoo.it](mailto:plucix@yahoo.it)

Relatore: Roberto Livi [liv@fi.infn.it](mailto:liv@fi.infn.it)

Lo scopo di questa tesi è affrontare lo studio delle proprietà strutturali di particolari sequenze di DNA, i promotori.

Queste sequenze hanno una funzione fondamentale nella regolazione dell'espressione genica. Nel DNA, in particolare nei geni, è codificata tutta l'informazione che determina le caratteristiche e lo sviluppo di un organismo vivente; ma questa informazione necessita di meccanismi che specificano quando e in quali casi deve essere utilizzata. Quindi, nel DNA, oltre all'informazione contenuta nei geni, sono codificate - per esempio nei promotori - anche le istruzioni che ne permettono un corretto utilizzo.

La sequenza del promotore, una volta rappresentata come successione di simboli, si presta a essere studiata con metodi statistici e della teoria dell'informazione; in questa tesi, prendendo spunto da precedenti risultati che evidenziavano una correlazione tra la funzione del promotore e la sua composizione, sono stati messi a punto due metodi spettrali per caratterizzare principalmente i promotori della specie *Homo sapiens*. In ogni caso, tali metodi sono applicabili a qualunque altra specie il cui genoma sia stato sequenziato.

In primo luogo, si è proceduto a raggruppare i promotori di una data specie in gruppi dalle proprietà di composizione simili e in secondo luogo ad individuare particolari motivi (sottosequenze) distintivi di ciascun gruppo, corrispondenti a funzioni note, in grado di evidenziare una correlazione tra proprietà strutturali e funzione dei promotori.

Il primo obiettivo è stato raggiunto utilizzando un metodo di clustering spettrale che, in base a un criterio che stima la similitudine tra le sequenze (con opportuni algoritmi di allineamento), individua una suddivisione dei promotori di una data specie in classi di equivalenza, utilizzando un opportuno algoritmo basato sul metodo detto "k-means". Questo metodo ha il vantaggio di ottenere una suddivisione del campione analizzato senza richiedere nessuna informazione a priori sui promotori.

I risultati originali ottenuti per *Homo sapiens* restituiscono una suddivisione in 4 cluster dalle caratteristiche di composizione distinte, che con i metodi precedentemente applicati non potevano essere osservati. A scopo di confronto, lo stesso metodo è stato applicato, con qualche adattamento, anche ai promotori di altre due specie, un pesce (*Danio rerio*) e una pianta (*Arabidopsis thaliana*). Anche in questi due casi è stato possibile distinguere cluster di differente composizione. In particolare, l'analisi dei promotori della pianta rivela delle strategie di regolazione diverse da quelle che sembrano accomunare *H. sapiens* e *D. rerio*. Questi risultati confermano la possibilità di uno studio più sistematico di altre specie, per evidenziare possibili tendenze evolutive nella composizione e nella struttura dei promotori, anche alla luce di studi già esistenti.

Il secondo obiettivo è stato ottenuto facendo uso di un metodo che permette di individuare eventuali sottosequenze caratteristiche di ciascun gruppo.

E' stato osservato in letteratura, e confermato in questa tesi, che un disordine non omogeneo contraddistingue le sequenze dei promotori, nei quali sottosequenze a composizione omogenea o periodica sono intervallate da zone disordinate.

Per individuare le zone regolari all'interno di ciascun promotore, sono state sfruttate le proprietà di localizzazione dei modi vibrazionali della catena in presenza di disordine, utilizzando un opportuno modello Hamiltoniano (il modello di Peyrard-Bishop) per modellizzare il DNA a doppio filamento. I modi vibrazionali hanno permesso di mettere a punto un metodo che permette di individuare sottosequenze regolari.

Queste sottosequenze hanno consentito di calcolare un'entropia di partizione del singolo promotore, e di definire alcuni indicatori che danno risultati consistenti con il metodo del clustering.

Riprendendo in considerazione i quattro gruppi di promotori ottenuti dalla suddivisione in cluster, è stato verificato che ciascun gruppo è caratterizzato da sottosequenze regolari, mentre alcune di queste risultano comuni a tutti i cluster. Inoltre, si è osservato che un meccanismo di selezione favorisce la presenza di sottosequenze a composizione omogenea, mentre sono più rare quelle in cui nucleotidi diversi si alternano.

Questa analisi ha permesso anche di ottenere informazioni più precise sulle caratteristiche di ciascun cluster. In particolare, emerge che due dei quattro cluster la cui composizione non evidenziava particolari proprietà distintive, sono invece caratterizzati dalla presenza di sottosequenze regolari più lunghe degli altri, cosa che suggerisce un ruolo funzionale delle sottosequenze individuate, e di conseguenza, dei promotori che le contengono. Anche la posizione preferenziale lungo il promotore, in alcuni casi, è correlata con il cluster di appartenenza. Inoltre, le sequenze meno espresse (che sono anche più variabili) sono molto probabilmente associate a motivi di interazione con fattori di trascrizione altamente specifici. Si è anche trovato un certo grado di corrispondenza tra alcune delle sottosequenze così individuate con la funzione dei promotori appartenenti ai vari cluster.

Al contrario le più comuni, a composizione omogenea, sono legate a meccanismi di regolazione che agiscono determinando le proprietà fisiche della catena del DNA, o che favoriscono il reclutamento di fattori di trascrizione.