



UNIVERSITÀ DEGLI STUDI DI MILANO

Facoltà di Scienze e Tecnologie
Laurea Triennale in Fisica

Studio di un modello stocastico per la trascrizione e l'esportazione dell'mRNA nel citoplasma

Relatore: Prof. Guido Tiana

Sara Terzoli
842044 n° 999999
A.A. 2016/2017

Codice PACS: 82.90.+j

Indice

1 Espressione genica	5
1.1 Trascrizione genica	5
1.2 Regolazione dell'espressione genica	7
2 Dati Sperimentali	10
3 Modello mRNA nascente	12
3.1 La trascrizione è un processo stocastico	12
3.2 Master equation	12
3.3 Risoluzione master equation	13
3.4 Momenti della distribuzione di probabilità	15
4 Studio dei rate di trascrizione	18
4.1 Andamento rate	18
4.2 Andamento degli errori dei rate	25
5 Distribuzione di probabilità del citoplasma	29
A Processo di Markov	35
A.1 Ricavazione Master Equation	36
A.2 Processo one-step	37
B Processo Random-Telegraph	37
C Funzione generatrice	37
D Bootstrap	38
E Espressioni analitiche rate	39
Bibliografia	43

Indice

Introduzione

In questo lavoro di tesi ci si occuperà di interpretare il processo di trascrizione genica con un approccio probabilistico. In particolare verrà studiato un modello stocastico a due stati che prenda in considerazione sia i processi che avvengono nel nucleo che quelli nel citoplasma; i due stati corrispondono a una situazione in cui il promotore è attivo e si verifica la produzione di mRNA e uno in cui è spento e non si ha produzione. Il modello teorico verrà sviluppato mediante la teoria delle master equation nonchè utilizzato per dedurre le equazioni dei rate che regolano il processo trascrizionale. Nello specifico i rate saranno espressi in funzione dei primi due momenti delle distribuzioni stazionarie di probabilità di nucleo e citoplasma. Le distribuzioni sperimentali riportate nell'articolo [1], a cui si farà riferimento per ricavare i momenti, sono state determinate sperimentalmente con la tecnica dell' smFISH.

Il presente lavoro risulta pertanto organizzato in cinque capitoli.

Dopo una breve descrizione biologica del processo trascrizionale (cap.I) e un' introduzione della tecnica sperimentale con cui sono state ricavate le distribuzioni sperimentali della probabilità dell' mRNA nucleare e citoplasmatico (capitolo II), si è passati alla modellizzazione del processo stesso e alla determinazione dell'espressioni analitiche per i rate (capitolo III). Nel capitolo IV, a partire dalle distribuzioni sperimentali di probabilità di nucleo e citoplasma in figura 7 determinate al variare del ciclo cellulare, verranno determinati i momenti di nucleo e citoplasma e quindi i rate e ne verranno studiati gli andamenti durante il ciclo cellulare. Nel capitolo V verrà ricostruita la distribuzione di probabilità del citoplasma a partire dai primi tre momenti di nucleo e citoplasma. Nel capitolo VI verranno ricapitolati i principali risultati ottenuti e tratte le conclusioni.

1. Espressione genica

1 Espressione genica

L'espressione genica è l'insieme dei processi attraverso il quale l'informazione contenuta in un gene viene convertita in una macromolecola funzionale, tipicamente in una proteina (figura 1).

La trascrizione e la traduzione sono i due stadi fondamentali che conducono alla sintesi delle proteine. Nel primo stadio, la molecola di DNA funge da stampo per la sintesi dell' RNA che è catalizzata dall' RNA polimerasi. Nel secondo stadio l'informazione genica contenuta nell'mRNA, viene convertita nella specifica proteina dai ribosomi.

Nelle cellule eucariotiche i due meccanismi appena descritti avvengono in compartimenti separati della cellula, la trascrizione nel nucleo e la traduzione nel citoplasma, mentre nelle cellule procariotiche, poichè non presentano alcuna membrana nucleare che delimiti il nucleo, avvengono entrambi nel citoplasma.

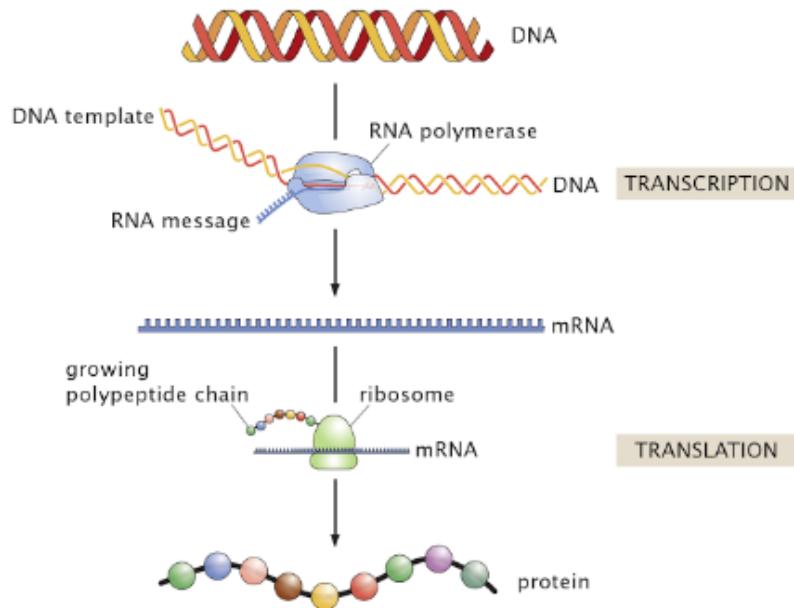


Figura 1: **Il dogma centrale.** Il filamento di DNA viene trascritto in uno di RNA che, a sua volta, porta l'informazione per la biosintesi delle proteine [6].

1.1 Trascrizione genica

Con il termine trascrizione si indica il processo, mediante il quale, la sequenza nucleotidica di un gene viene copiata in una molecola complementare di RNA,

1. Espressione genica

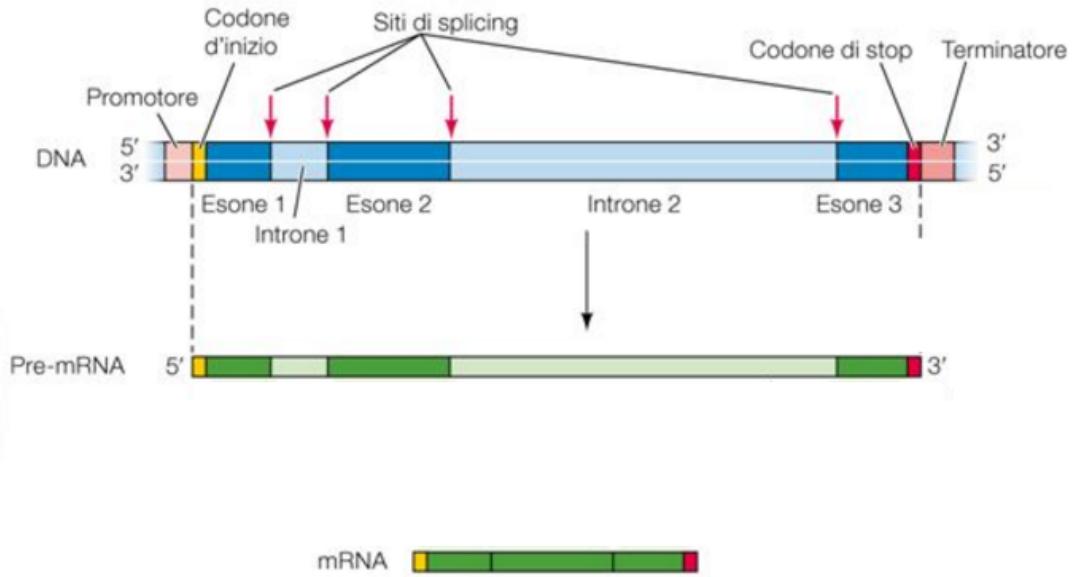


Figura 2: **Maturazione mRNA.** Il pre-mRNA, prima di essere tradotto, deve andare incontro a maturazione mediante la rimozione degli introni.

un polimero lineare a singolo filamento costituito da catene di ribonucleotidi. La trascrizione per essere attivata richiede che l'enzima RNA polimerasi si leggi al DNA in corrispondenza del *promotore*, una regione costituita da sequenze specifiche di nucleotidi posizionate a monte del punto di inizio della sintesi dell'mRNA. Una volta legata al DNA, l'RNA polimerasi catalizza la formazione di legami fosfodiesterici tra i ribonucleotidi che andranno a formare la catena di mRNA. L'enzima procede lungo il filamento di DNA in direzione 5'-3', aggiungendo nucleotidi secondo la regola dell'accoppiamento delle basi, il filamento di DNA utilizzato come stampo sarà quindi quello in direzione 3'-5'. Quando l'RNA polimerasi incontra una particolare sequenza di nucleotidi, il *terminatore*, rilascia sia il filamento di DNA che il *pre-mRNA* e termina così la trascrizione.

I meccanismi appena descritti sono simili nelle cellule eucariote e in quelle procariote, le differenze sostanziali tra i due tipi di cellule riguardano le fasi di maturazione e di esportazione del pre-mRNA. L'mRNA procariota infatti è già maturo dopo la trascrizione mentre quello eucariota deve andare incontro allo *splicing*. Vengono rimosse le sequenze non codificanti, gli *introni*, si forma così l'mRNA maturo costituito solo da sequenze codificanti, gli *esoni*. L'mRNA procariotico non necessita di essere trasportato fuori dal nucleo e la sintesi proteica può iniziare immediatamente, quello eucariota invece viene trasportato al citoplasma dove

1. Espressione genica

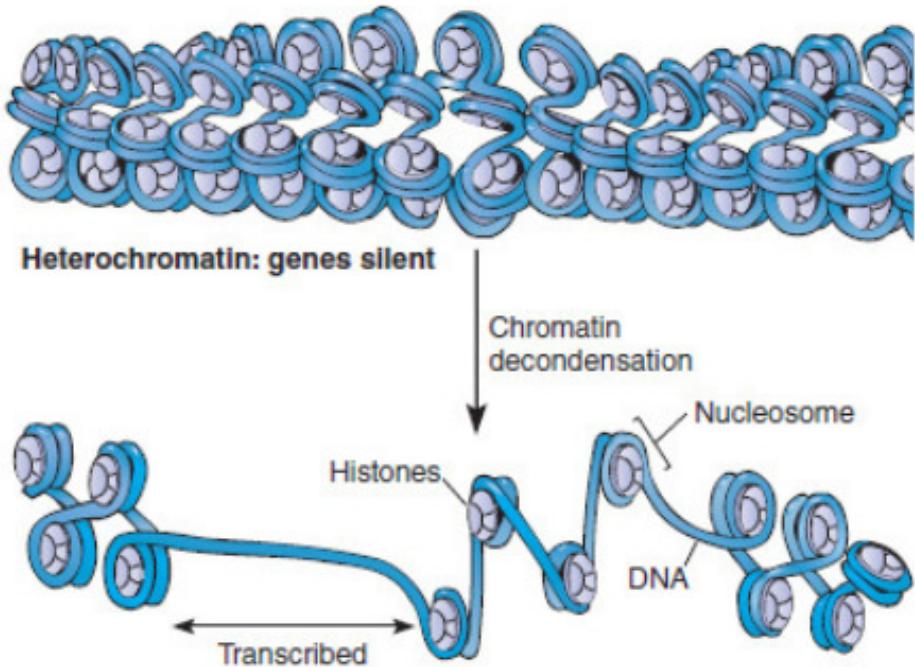


Figura 3: **Condensazione della cromatina.** Struttura della chromatina nelle cellule eucariote.

verrà tradotto dal ribosoma per codificare la specifica proteina [6].

1.2 Regolazione dell'espressione genica

La regolazione dell'espressione genica negli organismi è alla base del differenziamento cellulare, le cellule somatiche di una specie pur condividendo lo stesso patrimonio genetico lo esprimono in modo differente in relazione alla funzione che svolgono.

L'espressione genica viene regolata in più stadi del processo di sintesi proteica. Nei procarioti il controllo avviene prevalentemente a livello della trascrizione ed è basato sul modello dell'operone. L'operone è un complesso costituito da uno o più geni strutturali che codificano per determinati enzimi, da un promotore a cui si lega l'RNA polimerasi e da un operatore situato a monte o a valle della trascrizione a cui si lega una particolare proteina chiamata proteina repressore o attivatore. Quando un repressore è legato all'operatore, il promotore viene ostacolato e quindi l'RNAP non può legarsi alla molecola di DNA.

Negli eucarioti i meccanismi di controllo sono più complessi e coinvolgono anche la fase di maturazione dell' mRNA, la fase di trasporto nel citoplasma e quella di sintesi. I più importanti meccanismi di regolazione riguardano lo stato di con-

1. Espressione genica

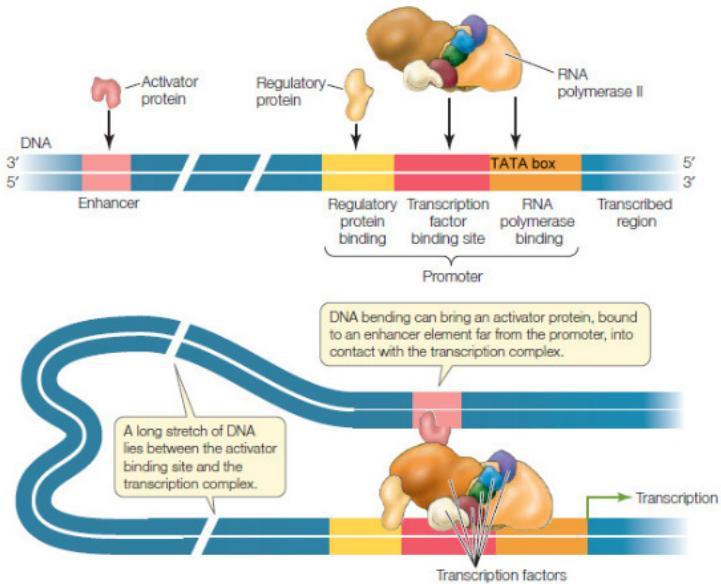


Figura 4: **Complesso di trascrizione.** Interazione enhancer-promotore mediata dall'RNA polimerasi e dal rimodellamento della cromatina.

densazione della cromatina, l'attività dei fattori di trascrizione e delle proteine regolatrici e la presenza dell' enhancer [8].

Cromatina

Il DNA negli eucarioti non è libero ma è strettamente associato agli *istoni*, proteine strutturali, e a proteine cromosomiche non istoniche, il complesso costituito da istoni e DNA prende il nome di cromatina, la cui unità fondamentale è il *nucleosoma*. Gli istoni sono proteine relativamente piccole che contengono un'elevata quantità di amminoacidi carichi positivamente, la lisina e l' arginina, che favoriscono la formazione di uno stretto legame con il DNA. La stretta associazione che vincola il DNA agli istoni lo rende inaccessibile alla RNAP (figura 3)

Per permettere la trascrizione è necessario che i residui di lisina vengano acetilati in modo da allentare la struttura della cromatina, mediante l'acetilazione viene infatti rimossa la carica positiva di questo amminoacido, di conseguenza, il legame tra il DNA che è carico negativamente, e le proteine istoniche sarà più debole.

1. Espressione genica

Fattori di trascrizione e proteine regolatrici

L'RNAP non inizia autonomamente la trascrizione ma ha bisogno dell'ausilio dei fattori di trascrizione, l'intero complesso prende il nome di complesso basale di trascrizione (figura 4). Il compito di queste proteine è di segnalare alla RNA Polimerasi II la collocazione del promotore su un dato gene, stabilizzarla in loco, aprire la doppia elica di DNA, attivare la RNA Polimerasi II e svolgere tutte quelle funzioni che garantiscono un corretto inizio del processo di trascrizione.

Enhancer

Gli enhancer sono sequenze nucleotidiche, lunghe in media 100pb, poste a parecchie centinaia di migliaia di paia di basi di distanza, a valle o a monte, del sito d'inizio della trascrizione, alle quali si legano proteine attivatrici che stimolano ulteriormente il complesso di trascrizione. Il contatto tra enhancer e promoter avviene in seguito al rimodellamento della cromatina (figura4).

2 Dati Sperimentali

In questa sezione vengono presentati i dati sperimentali, utilizzati da *Skinner et al.* [1]; i quali misurarano contemporaneamente il numero di molecole di mRNA nascente e maturo per cellula allo stato stazionario e determinarono quindi le distribuzioni di probabilità sperimentali in figura 6. I dati sperimentali a cui si farà riferimento in questo lavoro di tesi, verranno estratti dalle distribuzioni di probabilità presenti nell'articolo [1].

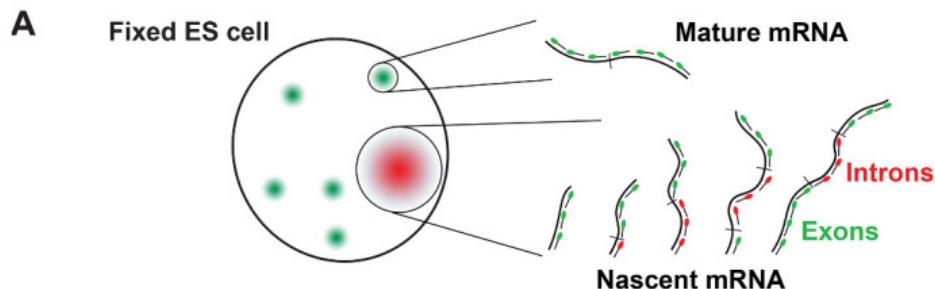


Figura 5: **Mrna nascente e maturo** In verde sono marcati gli esoni mentre in rosso gli introni. Nell' mRNA nascente si nota la presenza sia di introni che di esoni, mentre in quello maturo sono presenti solo gli esoni.[1]

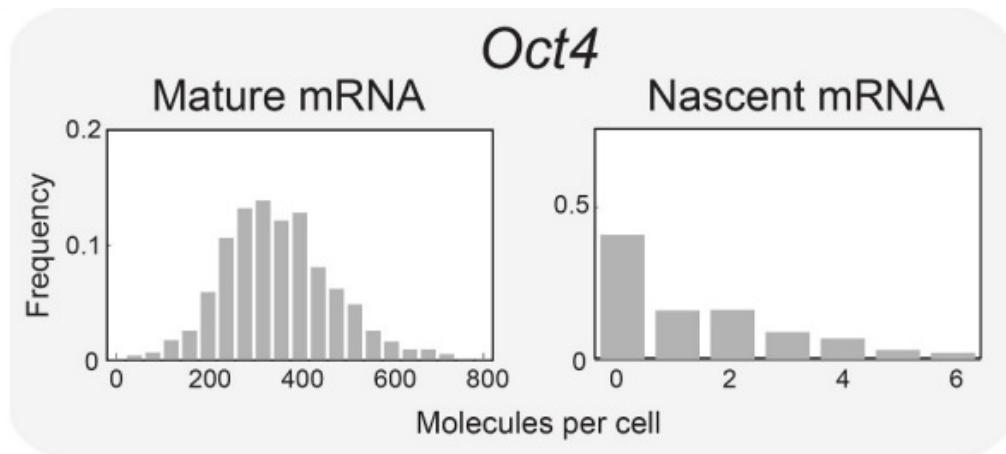


Figura 6: **Conteggi mRNA** Grafici dei conteggi di mRNA maturo e nascente per cellula nel gene Oct4 [1].

Skinner et al. studiarono la dinamica del processo trascrizionale misurando il numero di molecole di Mrna presenti nelle singole cellule [1].

2. Dati sperimentali

Gli esperimenti vennero condotti su geni, Oct4 e Nanag, presenti nel DNA dei topi e responsabili della regolazione della trasformazione delle cellule staminali embrionali in altri tipi di cellule.

Furono i primi a misurare simultaneamente il numero di molecole di mRNA nucleari e citoplasmatiche per cellula. Il trascritto primario, presente nel nucleo, prende il nome di mRNA nascente ed è costituito sia da sequenze codificanti, gli esoni, che non codificanti gli introni; l'mRNA citoplasmatico, costituito solamente da sequenze codificanti, prende il nome di mRNA maturo. L' mRNA nascente viene trasferito nel citoplasma in seguito alla crescita e allo splicing per poi essere degradato in seguito alla traduzione.

La tecnica con cui furono effettuate queste misurazioni è quella dell'smFISH (single-molecule fluorescence in situ hybridization): vengono utilizzate due sonde a fluorescenza per marcare le sequenze di esoni e quelle di introni. Le sequenze di introni sono marcate in rosso mentre quelle degli esoni in verde. Viene poi svolta un'analisi automatizzata delle immagini per individuare gli spots di esoni, introni e per identificare il contorno delle cellule. L'mRNA nascente sarà circondato da spots provenienti da entrambe le sonde, mentre l'mRNA maturo presenterà spots provenienti unicamente dalla sonda che marca gli esoni 5.

È stata successivamente identificata l'intensità di fluorescenza corrispondente a una singola molecola di Mrna maturo ed è stata utilizzata questa quantità per convertire l'intensità totale di fluorescenza degli spots degli esoni, nel numero di molecole di Mrna nascente e maturo.

Si possono costruire quindi gli histogrammi con i conteggi di mRNA nascente e maturo per cellula, per il gene Oct4, allo stato stazionario (figura 6).

A partire dalle distribuzioni sperimentali della quantità di mRNA nascente e maturo presenti nell'articolo, verranno successivamente ricavati i momenti della distribuzione di probabilità del nucleo $\langle m' \rangle, \langle m'^2 \rangle$ e del citoplasma $\langle m \rangle, \langle m^2 \rangle$ necessari per ricavare i rate di trascrizione.

3 Modello mRNA nascente

In questo capitolo viene presentata una modellizzazione del processo di trascrizione genica, con lo scopo di determinare analiticamente le espressioni dei rate che regolano il processo stesso, allo stato stazionario, in funzione dei momenti della distribuzione di probabilità, determinati a partire dalle distribuzioni sperimentali (figura 6).

3.1 La trascrizione è un processo stocastico

La trascrizione è un processo altamente stocastico, come conseguenza, cellule simili presentano livelli differenti di espressione genica: il numero di proteine sintetizzate varia molto da cellula a cellula, anche nella stessa popolazione. Questa variabilità è da imputare soprattutto al fatto che il processo di trascrizione non è un processo continuo ma avviene in *burst*; il promotore alterna casualmente due stati : uno in cui è attivo e avviene la trascrizione e l'altro in cui è spento e non viene prodotto mRNA.

Risulta complicato cercare di comprendere l'intrinseca stocasticità del processo stesso misurando la quantità di mRNA cellulare, poichè questo numero è influenzato da differenti fattori: la degradazione dell'mRNA nel citoplasma, la divisione all'interno del ciclo cellulare e la presenza di più copie dello stesso gene.

L'mRNA nascente al contrario non è soggetto a questi effetti essendo piú prossimo alla trascrizione, pertanto riflette meglio le caratteristiche stocastiche intrinseche del processo stesso [4].

Per ovviare questo problema, nel seguente capitolo viene migliorato il modello a due stati ipotizzato da Walczak [3], combinandolo con il modello per l'mRNA nascente.

Dalle equazioni che regolano il processo trascrizionale modellizzato in figura 8, si riescono a determinare le espressioni dei rate in funzione dei primi due momenti della distribuzione di probabilità dell'RNA nascente e maturo, ricavati dalle distribuzioni sperimentali di probabilità presenti nell'articolo di Skinner *et al.*[4] 6.

3.2 Master equation

Il modello schematizzato in figura 8, è costituito da quattro step: l'attivazione del gene, l'inizio della trascrizione con la sintesi dell'mRNA nascente, la conversione in mRNA maturo e la degradazione.

Si assume che il promotore si possa trovare in due stati e che vari stocasticamente da stato attivo a inattivo con rate di transizione $w+$ e da inattivo ad

2. Modello mRNA nascente

attivo con rate w_- . Quando il gene si trova nello stato inattivo non avviene trascrizione e non viene prodotto mRNA nucleare, quando invece si trova nello stato attivo inizia la trascrizione con rate costante g e viene prodotto mRNA nucleare. L'RNA nascente viene poi convertito in mRNA maturo con rate costante r e in seguito degradato con rate costante γ .

Il modello è sviluppato da un punto di vista matematico mediante le master equation tipiche del processo one step in appendice A.2 con l'aggiunta di termini necessari per descrivere il comportamento a due stati del promotore, cioè la probabilità di transizione dallo stato attivo a quello inattivo e viceversa in appendice B. Le master equation sono equazioni necessarie per descrivere l'evoluzione temporale della distribuzione di probabilità sul set di differenti stati in appendice A. Nel processo in esame la master equation descrive la variazione nel tempo della distribuzione di probabilità $p_{m',m}$ del numero di copie di mRNA nascente e mRNA maturo, indicati rispettivamente con la variabile discreta m' e m .

In particolare è stata ricavata la seguente master equation per lo stato attivo:

$$\begin{aligned}\dot{p}_{m',m}^+ = & +gp_{m'-1,m}^+ - gp_{m',m}^+ + r(m'+1)p_{m'+1,m-1}^+ - rm'p_{m',m}^+ \\ & + \gamma(m+1)p_{m',m+1}^+ - \gamma mp_{m',m}^+ + w_+p_{m',m}^- - w_-p_{m',m}^+\end{aligned}\quad (3.1)$$

E per lo stato inattivo :

$$\begin{aligned}\dot{p}_{m',m}^- = & +r(m'+1)p_{m'+1,m-1}^- - rm'p_{m',m}^- + \gamma(m+1)p_{m',m+1}^- \\ & - \gamma mp_{m',m}^- + w_-p_{m',m}^+ - w_+p_{m',m}^-\end{aligned}\quad (3.2)$$

Esse costituiscono un sistema di due equazioni differenziali accoppiate alle derivate parziali. L'equazione marginalizzata sulla variabile π che assume i valori 0 e 1 a seconda dello stato del promotore risulta

$$\begin{aligned}\dot{p}_{m',m} = & +gp_{m'-1,m}^+ - gp_{m',m}^+ + r(m'+1)p_{m'+1,m-1}^+ \\ & + r(m'+1)p_{m'+1,m-1}^- - rm'p_{m',m}^+ - rm'p_{m',m}^- + \gamma(m+1)p_{m',m+1}^+ \\ & + \gamma(m+1)p_{m',m+1}^- - \gamma mp_{m',m}^+ - \gamma mp_{m',m}^-\end{aligned}\quad (3.3)$$

3.3 Risoluzione master equation

L'approccio tipico per la risoluzione delle master equation consiste nell' introduzione della funzione generatrice: attraverso le sue derivate si può infatti determinare la distribuzione di probabilità e i momenti. Nello specifico la funzione generatrice è definita nel seguente modo:

$$G(y, x, t) = \sum_{m'm}^{\infty} P_{m'm}(t) y^{m'} x^m \quad (3.4)$$

2. Modello mRNA nascente

dove la variabile y è riferita al nucleo mentre la x al citoplasma.

I momenti vengono determinati per derivazione della stessa:

$$\langle m^l \rangle = [(x\partial_x)^l G(y, x, t)]|_{x=1} \quad (3.5)$$

$$\langle m'^l \rangle = [(y\partial_y)^l G(y, x, t)]|_{y=1} \quad (3.6)$$

La funzione generatrice permette di trasformare due set infiniti di equazioni differenziali (3.1) e (3.2), in un'unica equazione alle derivate parziali per ogni stato. Nello specifico, per determinare questa equazione si somma, sugli indici m' e m la master equation, allo stato stazionario, moltiplicata per $y^{m'} x^m$.

Allo stato stazionario, per lo stato attivo (3.1) si ricava:

$$\begin{aligned} 0 = & g \sum_{m'm} p_{m'-1,m}^+ y^{m'} x^m - g \sum_{m'm} p_{m',m}^+ y^{m'} x^m + r \sum_{m'm} (m' + 1) p_{m'+1,m-1}^+ y^{m'} x^m \\ & - rm' p_{m',m}^+ y^{m'} x^m + \gamma \sum_{m'm} (m + 1) p_{m',m+1}^+ y^{m'} x^m - \gamma \sum_{m'm} mp_{m',m}^+ y^{m'} x^m \\ & + w_+ \sum_{m'm} p_{m',m}^- y^{m'} x^m - w_- \sum_{m'm} p_{m',m}^+ y^{m'} x^m \end{aligned} \quad (3.7)$$

Da cui:

$$\begin{aligned} 0 = & gyG_+, -gG_+ + rx\partial_y G_+ - ry\partial_y G_+ + \gamma\partial_x G_+ \\ & - \gamma x\partial_x G_+ + w_+G_- - w_-G_+ \end{aligned} \quad (3.8)$$

Allo stato stazionario per lo stato inattivo (3.2) si ricava:

$$\begin{aligned} 0 = & r \sum_{m'm} (m' + 1) p_{m'+1,m-1}^- y^{m'} x^m - rm' p_{m',m}^- y^{m'} x^m + \gamma \sum_{m'm} (m + 1) p_{m',m+1}^- y^{m'} x^m \\ & - \gamma \sum_{m'm} mp_{m',m}^- y^{m'} x^m - w_+ \sum_{m'm} p_{m',m}^- y^{m'} x^m + w_- \sum_{m'm} p_{m',m}^+ y^{m'} x^m \end{aligned} \quad (3.9)$$

Da cui:

$$\begin{aligned} 0 = & rx\partial_y G_- - ry\partial_y G_- + \gamma\partial_x G_- - \gamma x\partial_x G_- \\ & - w_+G_- + w_-G_+ \end{aligned} \quad (3.10)$$

Per il calcolo completo si veda l'appendice (C). Si ricavi l'espressione analitica di G_+ dalla (3.8) e la si sostituisca nella (3.10), ottenendo la seguente equazione:

$$\begin{aligned} 0 = & \frac{(-w_- + \gamma(1-x)\partial_x - ry\partial_y + rx\partial_y - g + gy)(w_+ - (1-x)\gamma\partial_x + ry\partial_y - rx\partial_y)G_-}{w_-} \\ & + w_+G_- \end{aligned} \quad (3.11)$$

2. Modello mRNA nascente

Nello specifico non è stato possibile risolvere analiticamente l'equazione (3.11) per la funzione generatrice, pertanto i momenti della distribuzione di probabilità verranno determinati con un approccio differente presentato nel paragrafo seguente.

3.4 Momenti della distribuzione di probabilità

Non disponendo della soluzione dell'equazione (3.11), non è stato possibile determinare la forma analitica della funzione generatrice $G(y, x, t)$, ci si è limitati quindi al calcolo dei primi due momenti della distribuzione di probabilità allo stato stazionario, per il nucleo e per il citoplasma a partire dalla master equation (3.3).

L'equazione dinamica per il primo momento $\langle m \rangle$ è stata determinata a partire dall'equazione (3.3) moltiplicata per m e sommata su m', m e π .

$$\begin{aligned} \dot{\langle m \rangle} &= g \sum_{m'm} mp_{m'-1,m}^+ - g \sum_{m'm} mp_{m'-1,m}^+ \\ &+ r \sum_{\pi m'm} m(m'+1)p_{m'+1,m-1}^\pi - r \sum_{\pi m'm} mm'p_{m',m}^\pi \\ &+ \gamma \sum_{\pi m'm} m(m+1)p_{m',m+1}^\pi - \gamma \sum_{\pi m'm} m^2 p_{m',m}^\pi \end{aligned} \quad (3.12)$$

Il termine a destra dell'equazione rappresenta la derivata temporale della media. Nella parte sinistra dell'equazione, l'indice π assume il valore 0 quando il promotore è spento e 1 quando è acceso, mentre gli indici m' e m assumono valori discreti compresi tra 0 e ∞ . Nel calcolo che segue sono state ridefinite le variabili m e m' : $m \equiv m-1$ e $m' \equiv m'+1$ ed è stato posto il limite inferiore di tutte le somme uguale a 0.

Si ha:

$$\begin{aligned} \dot{\langle m \rangle} &= g \sum_{m'm} mp_{m',m}^+ - g \sum_{m'm} mp_{m'-1,m}^+ \\ &+ r \sum_{\pi m'm} m'(m+1)p_{m',m}^\pi - r \sum_{\pi m'm} mm'p_{m',m}^\pi \\ &+ \gamma \sum_{\pi m'm} m(m-1)p_{m',m}^\pi - \gamma \sum_{\pi m'm} m^2 p_{m',m}^\pi \\ &= r \langle m' \rangle - \gamma \langle m \rangle \end{aligned} \quad (3.13)$$

Con un procedimento analogo, moltiplicando la (3.3) per m' si ottiene l'equazione dinamica per il primo momento $\langle m' \rangle$:

$$\dot{\langle m' \rangle} = g \langle \pi \rangle - r \langle m' \rangle \quad (3.14)$$

2. Modello mRNA nascente

Le equazioni dinamiche per i momenti secondi , ricavati analogamente ai momenti primi, hanno le seguenti espressioni:

$$\begin{aligned}\dot{\langle m^2 \rangle} &= r \langle m' \rangle + 2r \langle mm' \rangle + \gamma \langle m^2 \rangle - 2\gamma \langle m^2 \rangle \\ \dot{\langle m'^2 \rangle} &= 2g \langle m' \rangle_+ + g \langle \pi \rangle + r \langle m' \rangle - 2r \langle m'^2 \rangle \\ \dot{\langle mm' \rangle} &= g \langle m \rangle_+ - r \langle m'^2 \rangle - r \langle m'm \rangle - r \langle m' \rangle - \gamma \langle m'm \rangle\end{aligned}\tag{3.15}$$

dove $\langle \pi \rangle$ rappresenta il rapporto dei rate relativi allo stato del promoter:

$$\dot{\langle \pi \rangle} = w_+ - w_+ \langle \pi \rangle - w_- \langle \pi \rangle\tag{3.16}$$

mentre $\langle m' \rangle_+$ é stato ricavato a partire dalla master equation per lo stato inattivo (3.2):

$$\langle m' \rangle_- = \frac{w_- \langle m' \rangle_+}{(r + w_+)}\tag{3.17}$$

A partire dalle equazioni (3.13),(3.14) e (3.15) sono state ricavate le espressioni analitiche dei momenti della distribuzione di probabilità nel caso stazionario:

$$\begin{aligned}\langle m \rangle &= \frac{g}{\gamma} \langle \pi \rangle \\ \langle m' \rangle &= \frac{g}{r} \langle \pi \rangle \\ \langle m'^2 \rangle &= \frac{1}{2r}(g \langle \pi \rangle + r \langle m' \rangle + 2g \langle m' \rangle_+) \\ \langle m'm \rangle &= \frac{g \langle m \rangle_+ + r \langle m'^2 \rangle - r \langle m' \rangle}{r + \gamma} \\ \langle m^2 \rangle &= \frac{1}{2\gamma}(r \langle m' \rangle + 2r \langle mm' \rangle + \gamma \langle m \rangle)\end{aligned}\tag{3.18}$$

Invertendo le equazioni (3.18), sono state infine determinate le espressioni dei rate w_+ , w_- , g , r e γ in funzione dei momenti $\langle m' \rangle$, $\langle m \rangle$, $\langle m'^2 \rangle$ e $\langle m^2 \rangle$, i cui valori sono noti sperimentalmente. Il valore del rate γ è stato posto uguale a 1, quindi i rate ottenuti risultano riscalati su quest'ultimo. Il calcolo dettagliato e le soluzioni delle equazioni sono riportati in appendice E.

In ultima analisi è stato calcolato il momento terzo $\langle m'^3 \rangle$:

$$\begin{aligned}\dot{\langle m'^3 \rangle} &= 3g \langle m' \rangle_+ - r \langle m' \rangle + g \langle \pi \rangle + 3g \langle m'^2 \rangle_+ + 3r \langle m'^2 \rangle \\ &\quad - 3r \langle m'^3 \rangle\end{aligned}$$

2. Modello mRNA nascente

Analogamente a quanto svolto in appendice E sono stati calcolati i valori medi $\langle m'^2 \rangle_+$ e $\langle m'^2 \rangle_-$:

$$\begin{aligned}\langle \dot{m'^2} \rangle_+ &= 2g \langle m' \rangle_+ + g \langle \pi \rangle - 2r \langle m'^2 \rangle_+ + r \langle m' \rangle_+ + w_+ \langle m'^2 \rangle_- \\ &\quad - w_- \langle m'^2 \rangle_+ \\ \langle m'^2 \rangle_- &= \frac{w_- \langle m'^2 \rangle_+}{(-2r - r + w_+)}\end{aligned}$$

ed è stato sostituito il momento $\langle m'^2 \rangle_+$ nell'equazione per $\langle \dot{m'^3} \rangle$ per determinare il momento terzo $\langle m'^3 \rangle$.

$$\langle m'^3 \rangle = \frac{3g \langle m' \rangle_+ - r \langle m' \rangle + g \langle \pi \rangle + 3g \langle m'^2 \rangle_+ + 3r \langle m'^2 \rangle}{3r} \quad (3.19)$$

4 Studio dei rate di trascrizione

Nella prima parte di questo capitolo vengono analizzate le distribuzioni sperimentali presenti nell'articolo di *Skinner et al.* in figura 7, con l'obiettivo di determinare i momenti della distribuzione di probabilità del nucleo e del citoplasma. I risultati ottenuti verranno utilizzati per determinare i rate di trascrizione e per studiarne gli andamenti al variare del tempo del ciclo cellulare.

tempo [min]	w_+/γ
80	1.5107
160	0.205263
240	0.0374572
320	0.332271
400	1.74167
480	1.46772
560	0.735007
640	0.716302
720	0.170722
800	0.368827

Tabella 1: Andamento del rate w_+/γ al variare del tempo del ciclo cellulare per il gene Oct4.

4.1 Andamento rate

Se il ciclo cellulare ha una durata in media di qualche ora, d'altra parte la dinamica del processo di trascrizione avviene su scale temporali più brevi, dell'ordine dei minuti. Dal momento che le scale temporali dei due processi sono differenti, è lecito determinare le espressioni analitiche dei rate risolvendo le equazioni (3.13), (3.15) allo stato stazionario.

Con lo scopo di determinare l'andamento dei rate trascrizionali durante il ciclo cellulare, sono state analizzate le distribuzioni sperimentali in figura 7 presenti nell'articolo [1] da cui sono stati determinati i momenti necessari per ricavare i valori dei rate.

I grafici in figura 7 rappresentano infatti le distribuzioni di probabilità della quantità di mRNA nascente e maturo determinate in dieci intervalli temporali, egualmente spaziati, del ciclo cellulare. Per ogni grafico si è assunto che l'intervalle temporale in cui è stata misurata la quantità di mRNA fosse così breve rispetto

3. Studio dei rate di trascrizione

tempo [min]	g/γ
80	349.919
160	341.406
240	378.37
320	384.575
400	423.758
480	436.94
560	455.245
640	471.946
720	443.285
800	499.24

Tabella 2: Andamento del rate g/γ al variare del tempo del ciclo cellulare per il gene Oct4.

alla durata del ciclo cellulare da considerare le distribuzioni stazionarie. Ogni coppia di istogrammi è stata analizzata mediante il programma *Plot Digitizer* che restituisce ascisse e ordinate dei punti (x_i, y_i) del grafico : x_i rappresenta la quantità di mRNA, y_i la frequenza con cui si presenta una certa quantità di mRNA. A partire dalle coppie di dati (x_i, y_i) , sono stati calcolati i momenti mediante le seguenti equazioni:

$$\begin{aligned} \langle x \rangle &= \sum_{i=i}^{n-1} x_i p(x_i) = \sum_{i=i}^{n-1} \frac{x_i y_i (x_{i+1} - x_i)}{Z} \\ \langle x^2 \rangle &= \sum_{i=i}^{n-1} x_i^2 p(x_i) = \sum_{i=i}^{n-1} \frac{x_i^2 y_i (x_{i+1} - x_i)}{Z} \\ \langle x^3 \rangle &= \sum_{i=i}^{n-1} x_i^3 p(x_i) = \sum_{i=i}^{n-1} \frac{x_i^3 y_i (x_{i+1} - x_i)}{Z} \end{aligned} \quad (4.1)$$

con $Z = \sum_{i=i}^{n-1} y_i (x_{i+1} - x_i)$ coefficiente di normalizzazione della probabilità.

Per ogni fase del ciclo cellulare è stato quindi possibile determinare i valori dei rate, sostituendo nelle equazioni (E.13), (E.9) ,(E.11) i valori dei primi due momenti. I risultati ottenuti sono riportati nelle tabelle 1, 2, 3 e 4.

Il momento terzo verrà utilizzato per confermare la validità del modello teorico.

Una volta determinati i valori dei rate per ogni istogramma in figura 7 è interessante studiarne l'andamento al variare del tempo del ciclo cellulare.

3. Studio dei rate di trascrizione

tempo [min]	r/γ
80	150
160	162.65
240	200.00
320	179.13
400	208.36
480	247.00
560	227.80
640	254.15
720	186.00
800	242.20

Tabella 3: Andamento dei rate r/ω al variare del tempo del ciclo cellulare per il gene Oct4.

Sono stati quindi plottati i valori dei rate g , r , w_+ e w_- riportati nelle tabelle 1, 2, 3 e 4 e sono stati ottenuti rispettivamente i grafici in figura 9, 10, 11e 12.

Si può notare come nei primi due grafici l'andamento è lineare; per quantificare in che misura un insieme di coppie di punti, nello specifico $(t_i, rate_i)$ sostiene l'ipotesi di una relazione lineare, si calcola il coefficiente di correlazione lineare r , definito nel seguente modo:

$$r = \frac{\sum_i (x_i - \langle x \rangle)(y_i - \langle y \rangle)}{\sqrt{\sum_i (x_i - \langle x \rangle)^2 \sum_i (y_i - \langle y \rangle)^2}} \quad (4.2)$$

È un numero compreso tra -1 e 1, se è \simeq a ± 1 allora i punti giacciono vicino a qualche retta, se è \simeq a 0 allora i punti non sono correlati. È stato calcolato il coefficiente di correlazione r per i grafici in figura 9, 10, 11 e 12 e sono stati ottenuti risultati riportati nelle tabelle 5, 6, 7 e 8.

Per stimare il livello di confidenza dell'andamento ottenuto, è stato effettuato il test del P-Value. Per ogni valore osservato r_o , si consideri la quantità $P_N(|r| \geq |r_o|)$, cioè la probabilità che N misure di due variabili non correlate diano un coefficiente r grande quanto r_o : se il coefficiente r_o è tale che la $P_N(|r| \geq |r_o|)$ risulta piccola rispetto a un valore di riferimento fissato, allora è improbabile che le variabili siano non correlate.

Per mancanza di dati sperimentali su cui svolgere il test, è stato necessario generare M set di N coppie $(t_i, rate_i)$ di variabili scorrelate ed è stato calcolato per ogni set di misure il coefficiente di correlazione r . Prendendo come ipotesi nulla che le due variabili fossero scorrelate, è stato calcolato il coefficiente di correlazione r_θ sulle misure sperimentali ed è stata calcolata la probabilità di ottenere

3. Studio dei rate di trascrizione

tempo [min]	w_-/γ
80	0.463401
160	0.035772
240	0.00675665
320	0.0764249
400	0.500507
480	0.439657
560	0.219535
640	0.189798
720	0.0239777
800	0.0650617

Tabella 4: Andamento dei rate w_-/γ al variare del tempo del ciclo cellulare per il gene Oct4.

r_{w+}	-0.18127	< 74%
----------	----------	-------

Tabella 5: coefficienti di correlazione rate w_+ e relativo P-value.

$P_N(|r| \geq |r_o|)$. I risultati ottenuti sono riportati nelle tabelle 5, 6, 7 e 8.

Si può notare che si ha una probabilità < dello 0.003 % di ottenere $|r_g| \geq 0.955$ quando N=10, quindi è alquanto improbabile che esse siano non correlate.

Per quanto riguarda il rate r , si ha una probabilità < dello 0.6 % di ottenere $|r_r| \geq 0.7283$, quindi anche in questo caso, si può trascurare l'ipotesi nulla che siano scorrelate.

Nel caso del rate w_+ si ha invece una probabilità circa del 75 % di ottenere $|r_{w+}| \geq -0.1812$ quindi è assolutamente possibile che la variabile temporale e il rate w_+ non siano correlate come traspare in figura 11, per w_- si ha una probabilità del 14.6% di ottenere un coefficiente di correlazione uguale a 0.2215 quindi non si può concludere che l'ipotesi nulla possa essere pienamente trascurata.

I rate g e r seguono pertanto un andamento lineare al variare del tempo, questo potrebbe essere interpretato come un meccanismo di regolazione durante l'avanzamento del ciclo cellulare: un aumento dei momenti della distribuzione di pro-

r_{w-}	0.221545	< 14.6%
----------	----------	---------

Tabella 6: coefficienti di correlazione rate w_- e relativo P-value.

3. Studio dei rate di trascrizione

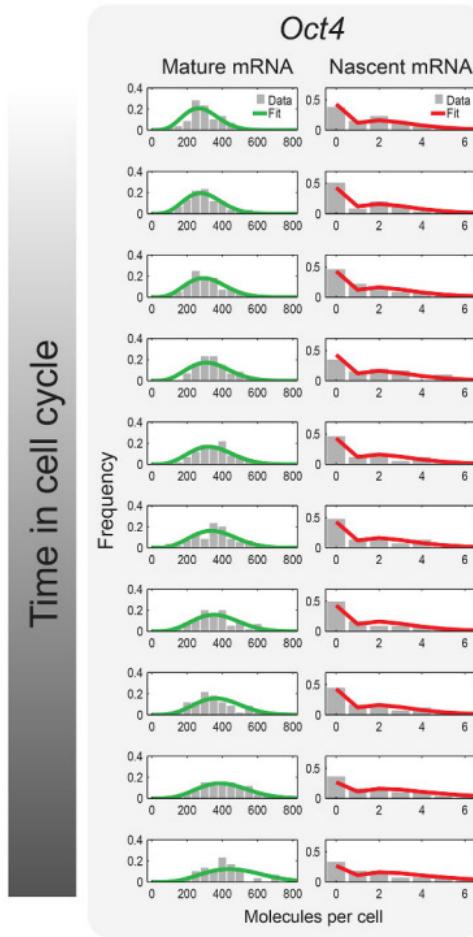


Figura 7: Distribuzione di probabilità Distribuzione di probabilità del numero di copie di mRNA maturo e nascente al variare delle fasi del ciclo cellulare per il gene Oct4 [1]. Le barre grigie indicano la quantità di molecole di mRNA per cellula, le linee verdi e rosse rappresentano rispettivamente la distribuzione di probabilità per l'mRNA nel citoplasma e nel nucleo, predetta dal modello teorico. I grafici sono stati ottenuti dividendo l'intervallo di tempo del ciclo cellulare (0-800 min) in 10 finestre temporali.

r_g	0.95525	$< 10^{-3}\%$
-------	---------	---------------

Tabella 7: coefficiente di correlazione rate g e relativo P-value.

3. Studio dei rate di trascrizione



Figura 8: **Schema a due stati del modello trascrizionale** In blu sono schematizzati i due stati del promotore: lo stato attivo in cui avviene la trascrizione con rate costante g e quello spento in cui non viene prodotto mRNA. L'mRNA nascente viene poi trasportato nel citoplasma con rate r e in seguito degradato con rate γ .

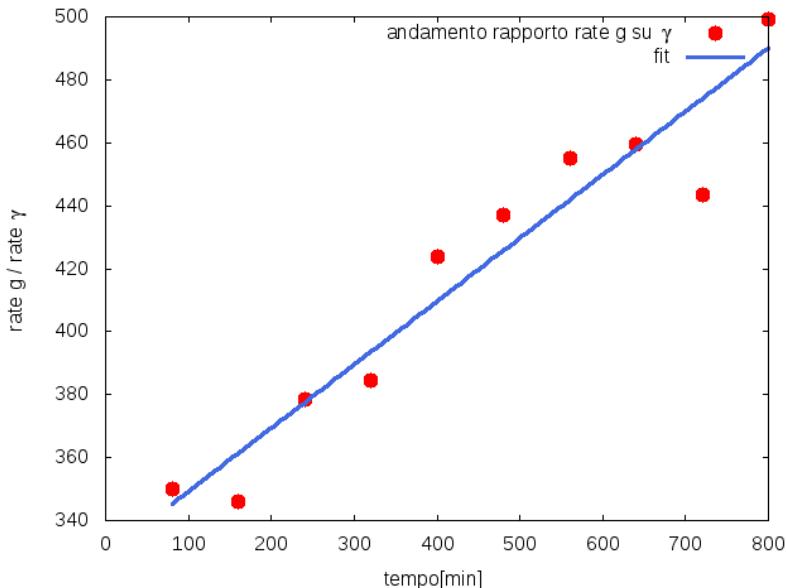


Figura 9: Andamento rate g su γ , al variare del tempo del ciclo cellulare.

babilità corrisponde a un aumento della velocità del processo trascrizionale. Il meccanismo di regolazione, che è stato supposto, avviene soprattutto a livello del nucleo dal momento che il rate g ha un andamento lineare più marcato rispetto al rate r .

Non si osserva lo stesso comportamento per quanto riguarda i rate w_+ e w_- , durante il ciclo cellulare assumono valori che oscillano tra lo 0.1 e lo 0.6 per il rate w_- e lo 0.1 e 2 per w_+ senza avere tuttavia un andamento definito.

Il comportamento del promotore può essere modellizzato durante tutta la fase cellulare con il processo random telegraph B: esso alterna stocasticamente lo stato in cui è attivo e lo stato in cui è spento.

Per avere una conferma della validità del modello studiato, abbiamo sostituito i

3. Studio dei rate di trascrizione

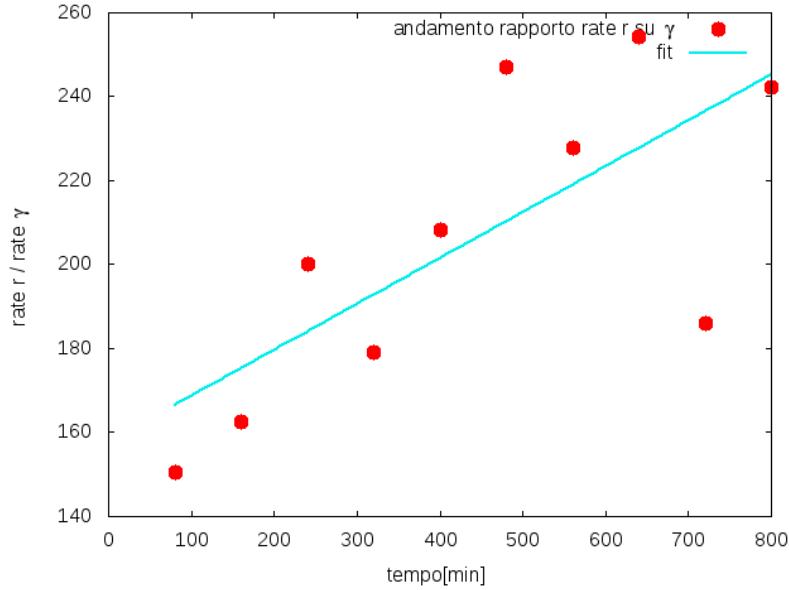


Figura 10: Andamento rate r su γ , al variare del tempo del ciclo cellulare.

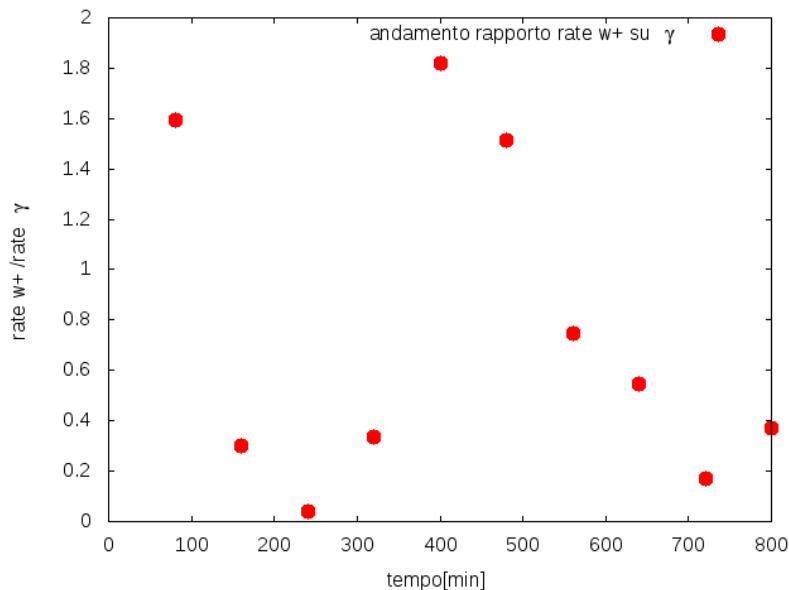


Figura 11: Andamento rate w_+ su γ , al variare del tempo del ciclo cellulare.

r_r	0.7283	0.6%

Tabella 8: coefficienti di correlazione rate r e relativo P-value.

3. Studio dei rate di trascrizione

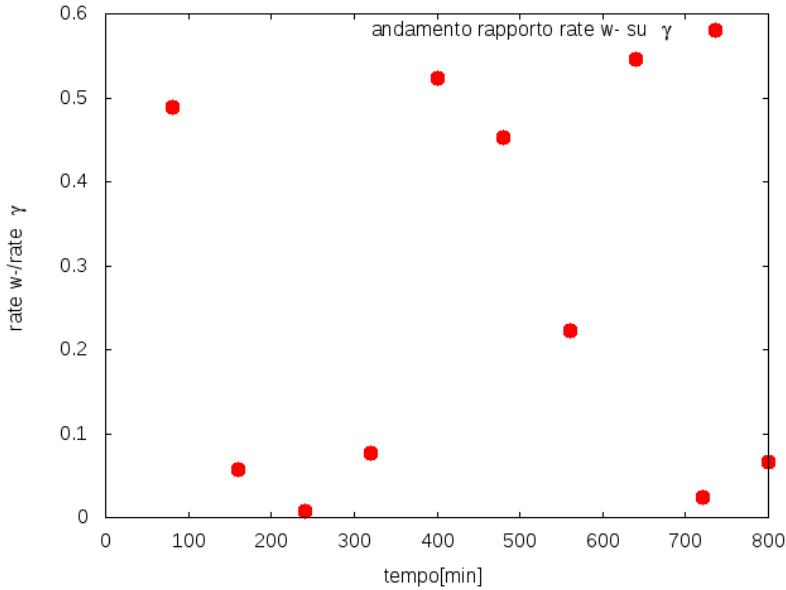


Figura 12: Andamento rate w_- su γ , al variare del tempo del ciclo cellulare.

valori dei rate ricavati dalla prima coppia di istogrammi di figura 7 nell'equazione (3.19) e abbiamo verificato l'identità con il momento terzo determinato sperimentalmente con la (4.1).

4.2 Andamento degli errori dei rate

Per mancanza di dati con cui stimare gli errori sui rate g , r , w_+ e w_- , è stato necessario simulare distribuzioni di dati a partire dalle distribuzioni sperimentali (seconda coppia di istogrammi in figura 7), assunte come distribuzioni vere. Sono stati quindi generati $N=1000$ istogrammi con il metodo *bootstrap* secondo la distribuzione di probabilità sperimentale del citoplasma e altri $N=1000$ istogrammi secondo la distribuzione sperimentale del nucleo. Per ognuno di essi sono stati calcolati, mediante l'equazione (4.1), i primi due momenti della distribuzione di probabilità.

Si assume che i momenti così determinati non abbiano errore. I momenti ottenuti sono stati sostituiti nelle equazioni (E.13) per determinare i rate.

Ad ogni rate determinato precedentemente dalla distribuzione sperimentale (seconda coppia di istogrammi in figura 7) è stato associato il relativo errore assoluto e percentuale determinato con metodi statistici. I risultati sono riportati in tabella 9. Dall'analisi degli errori sui rate emerge che è possibile determinare i rate

3. Studio dei rate di trascrizione

g	341,406	44,677	13 %
r	162,65	19,245	12 %
w_+	0,2052	2,69	130 %
w_-	0,0357	0,247	730 %
$\langle \pi \rangle$	0,85159	0,24484	29 %

Tabella 9: Valori rate di trascrizione con rispettivi errori assoluti e errori relativi percentuali.

g e r con una certa precisione, gli errori sono intorno al 12% - 13%; per quanto riguarda il termine $\langle \pi \rangle$ l'errore relativo percentuale è più basso rispetto all'errore sui singoli rate w_+ e w_- , questo suggerisce che è possibile determinare con precisione il comportamento del promotore, mentre non è possibile determinare singolarmente i rate di passaggio dallo stato attivo allo stato spento w_- e viceversa w_+ e . L'errore su $\langle \pi \rangle$ non viene determinato con i metodi di propagazione degli errori poichè le due variabili sono correlate.

Si considerino le seguenti espressioni ricavate nel capitolo precedente:

$$\begin{aligned} \langle m \rangle &= \frac{g}{\gamma} \langle \pi \rangle \\ \langle m' \rangle &= \frac{g}{r} \langle \pi \rangle \end{aligned} \quad (4.3)$$

Sono state graficate l'equazioni (4.3) in funzione rispettivamente dei rate g e π e sono stati ottenuti i grafici in figura 13, 14.

In entrambi i casi si nota che sussiste una relazione lineare tra i momenti della distribuzione di probabilità e i rate, con coefficiente angolare maggiore di 1 e come una variazione nei rate influenzi maggiormente le quantità legate al citoplasma rispetto a quelle del nucleo.

Per verificare che le distribuzioni di probabilità dell' mRNA nascente e maturo sono correlate, è stato calcolato il coefficiente di correlazione $\langle mm' \rangle - \langle m \rangle \langle m' \rangle$ per tutti i grafici in figura 7 a partire dalle conoscenze dei rate: sono stati ottenuti valori compresi tra 0.25 e 0.5.

Dalla prima equazione delle (E.13), si può notare infatti che intercorre una relazione lineare tra la quantità di mRNA del citoplasma $\langle m \rangle$ e la quantità di mRNA nel nucleo $\langle m' \rangle$, con coefficiente angolare dato dal rapporto tra il rate di trasferimento di mRNA tra nucleo e citoplasma e il rate di disintegrazione dell'mRNA nel citoplasma: dal primo momento della distribuzione di probabilità del nucleo e dalla conoscenza dei rate si possono ricavare informazioni sulla quantità

3. Studio dei rate di trascrizione

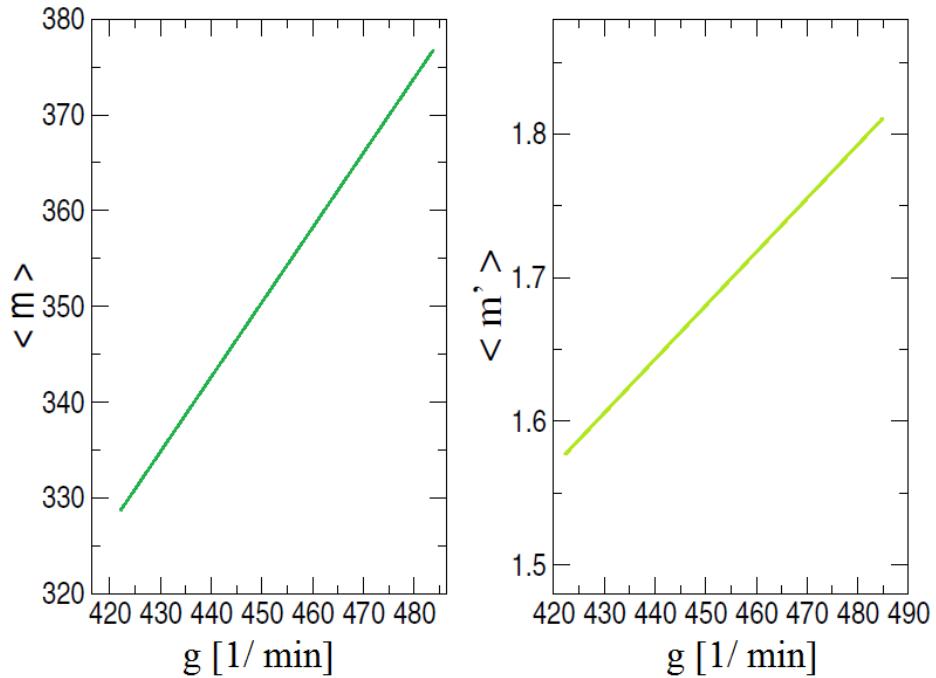


Figura 13: andamento $\langle m \rangle$ e $\langle m' \rangle$ al variare del rate g

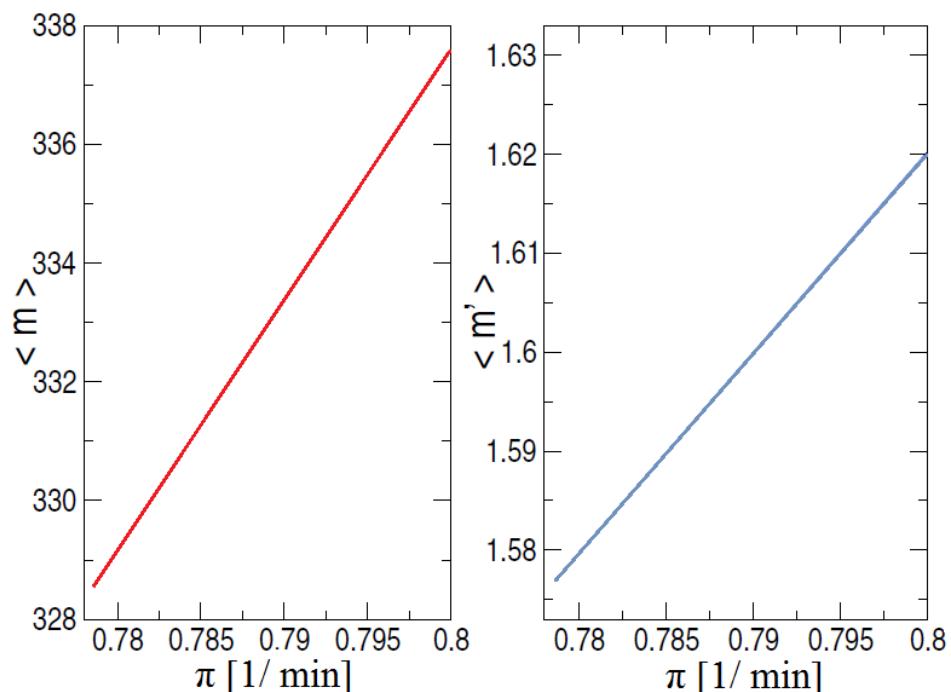


Figura 14: andamento $\langle m \rangle$ e $\langle m' \rangle$ al variare del rate π .

3. Studio dei rate di trascrizione

di mRNA nel citoplasma.

Nel prossimo capitolo a partire dai primi tre momenti, si studierà la distribuzione di probabilità dell' mRNA nel citoplasma in modo da avere una visione più completa del fenomeno della trascrizione.

5 Distribuzione di probabilità del citoplasma

Come visto in precedenza, non è stato possibile determinare l'espressione analitica della funzione generatrice e quindi un'espressione per la distribuzione di probabilità $P_{m',m}$, dalla quale poter ottenere tutti i momenti.

Sono state quindi ricavate le espressioni analitiche per i primi tre momenti a partire dalla master equation (3.3).

Lo scopo di questo capitolo è dimostrare che la distribuzione di probabilità di nucleo e citoplasma determinata a partire almeno dai primi tre momenti della distribuzione di probabilità riesce a rappresentare la distribuzione di probabilità sperimentale in figura 6.

Si è scelto di studiare solo la distribuzione di probabilità del citoplasma tenendo presente che nel caso del nucleo il procedimento sarebbe stato analogo.

L'espressione analitica per la distribuzione di probabilità del citoplasma è stata ricavata sfruttando il principio di massima entropia [7]: la distribuzione che rappresenta meglio lo stato attuale del sistema è quella che rende l'entropia del sistema massima, poiché si hanno informazioni su i primi tre momenti della distribuzione di probabilità, sono stati posti 4 vincoli all'entropia di Shannon.

La distribuzione così ricavata ha la seguente forma:

$$P(x) = \frac{e^{-\lambda_1 x - \lambda_2 x^2 - \lambda_3 x^3}}{Z} \quad (5.1)$$

Con Z coefficiente di normalizzazione e $\lambda_1, \lambda_2, \lambda_3$ moltiplicatori di lagrange che compaiono per aver posto i vincoli all'entropia di Shannon.

I moltiplicatori di lagrange si ottengono per derivazione della funzione di partizione, nel seguente modo:

$$\begin{aligned} \langle m_k \rangle &= \frac{-\partial \log Z(\lambda_1..\lambda_m)}{\lambda_k} \\ &= \frac{-\partial \log \sum_{i=1}^n e^{-\sum_{k=1}^m \lambda_k x_i}}{\lambda_k} \end{aligned}$$

I valori numerici per i moltiplicatori di lagrange sono stati ricavati mediante il fit dei dati sperimentali del primo grafico in figura 7 poichè non è stato possibile determinarli per derivazione della funzione di partizione.

Il grafico in figura 15 mostra l'andamento sperimentale della distribuzione di probabilità di mRNA nel citoplasma fittata mediante la funzione (5.1).

In tabella 10 sono riportati i valori di $Z, \lambda_1, \lambda_2, \lambda_3$ ottenuti dal fit.

In seguito, la stessa distribuzione di probabilità sperimentale è stata fittata con

4. Distribuzione di probabilità del citoplasma

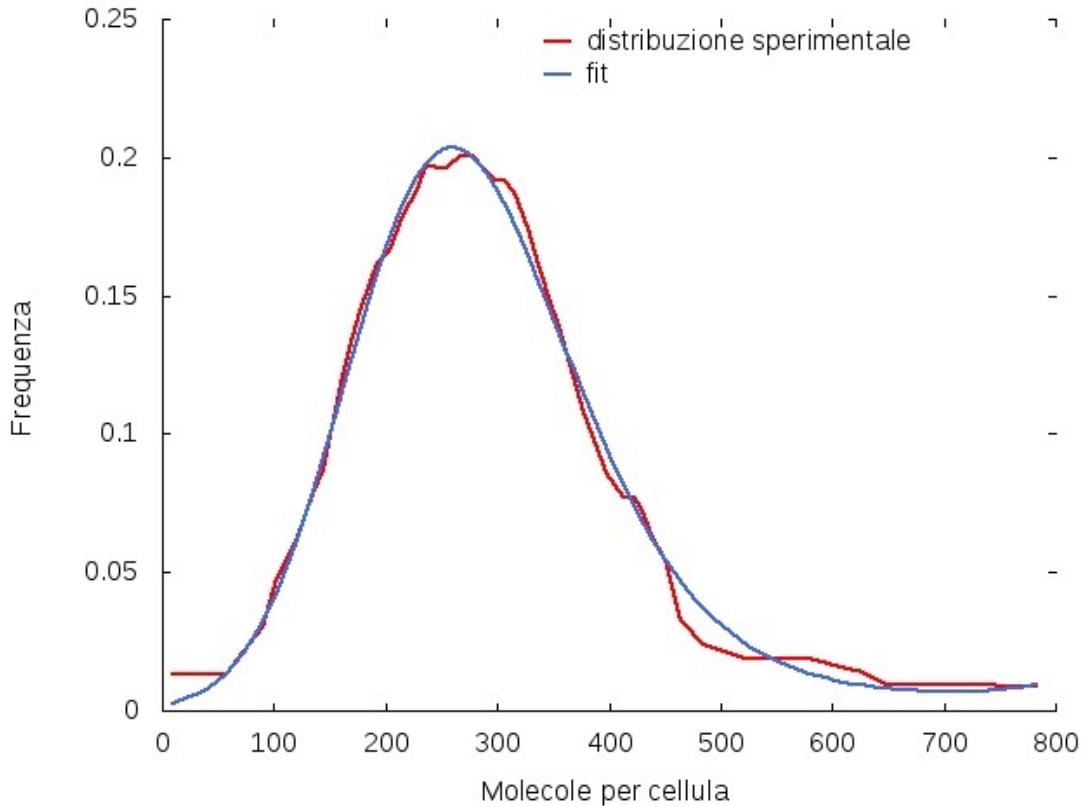


Figura 15: **Distribuzione di probabilità citoplasma** Curva dati sperimentali in rosso e il relativo fit in blu. Sulle ascisse è riportata la quantità di mRNA maturo per cellula misurata allo stato stazionario, sulle ordinate la frequenza.

una funzione gaussiana:

$$P(x) = \frac{\lambda_1 e^{-(x-\lambda_2)^2}}{\lambda_3^2} \quad (5.2)$$

Il grafico in figura 16 mostra l'andamento sperimentale della distribuzione di probabilità di mRNA nel citoplasma fittata mediante la funzione gaussiana.

In tabella 11 sono riportati i valori ottenuti dal fit.

Sono stati sostituiti nelle equazioni (5.1),(5.2) i valori ricavati dai fit e la quantità di mRNA x_i ricavata precedentemente dal primo grafico in figura 6, con $i=(1\dots N)$ ed N numero di punti del grafico considerati, così da determinare due set di valori per le ordinate $f(x_i)_{exp}, f(x_i)_{gauss}$.

Per quantificare i risultati ottenuti dai due fit è stato calcolato per i due fit il $\chi^2 = \sum_{i=1}^N \frac{y_i - f(x_i)}{f(x_i)}$.

I risultati ottenuti sono riportati in tabella 12.

Il fatto che il χ^2 calcolato utilizzando la gaussiana come funzione per il fit, abbia

4. Distribuzione di probabilità del citoplasma

λ_1	λ_2	λ_3	Z
-0.04	0.0001	-7.55E-008	573.053

Tabella 10: Valori fit 5.1.

λ_1	λ_2	λ_3
0.2	270	145

Tabella 11: Valori fit gaussiana.

un valore così alto, dimostra che la distribuzione di probabilità sperimentale non è determinata dalla conoscenza dei primi due momenti e non ha un andamento gaussiano, quindi per poterla fissare sarà necessario considerare almeno i primi tre momenti della distribuzione di probabilità.

Confrontando i due grafici in figura 15, 16, si può osservare che la funzione gaussiana non riesce a fissare le code della distribuzione di probabilità e per questo motivo il χ^2 ha un valore così alto.

Ancora una volta, per avere una misura più quantitativa del risultato ottenuto, è stato effettuato il test del *P-VALUE*.

Si considera come ipotesi nulla il fatto che i dati sperimentali non si distribuiscono secondo la distribuzione di probabilità (5.1); si calcola il χ^2_{gauss} per N distribuzioni gaussiane di dati sperimentali: se la $P_N(|\chi^2_{gauss}| \leq |\chi^2_{exp}|)$ è piccola, allora è altamente improbabile che i dati non siano distribuiti secondo la (5.1).

Per mancanza di dati sperimentali su cui svolgere il test, sono stati generati con il metodo *bootstrap* N set di dati distribuiti secondo la distribuzione di probabilità sperimentale riportata nel primo grafico in figura 7, assunta come distribuzione vera.

È stato ripercorso per ogni istogramma generato il procedimento descritto in precedenza: ogni istogramma è stato fissato con una distribuzione gaussiana ed è stato poi calcolato il χ^2 .

La probabilità di ottenere dei valori di χ^2 minori di quelli ottenuti ipotizzando che i dati seguano la distribuzione (5.1) è $\leq 10^{-3}$, perciò si può concludere che la

χ^2_{exp}	χ^2_{gauss}
0.098	334

Tabella 12: Valori fit gaussiana.

4. Distribuzione di probabilità del citoplasma

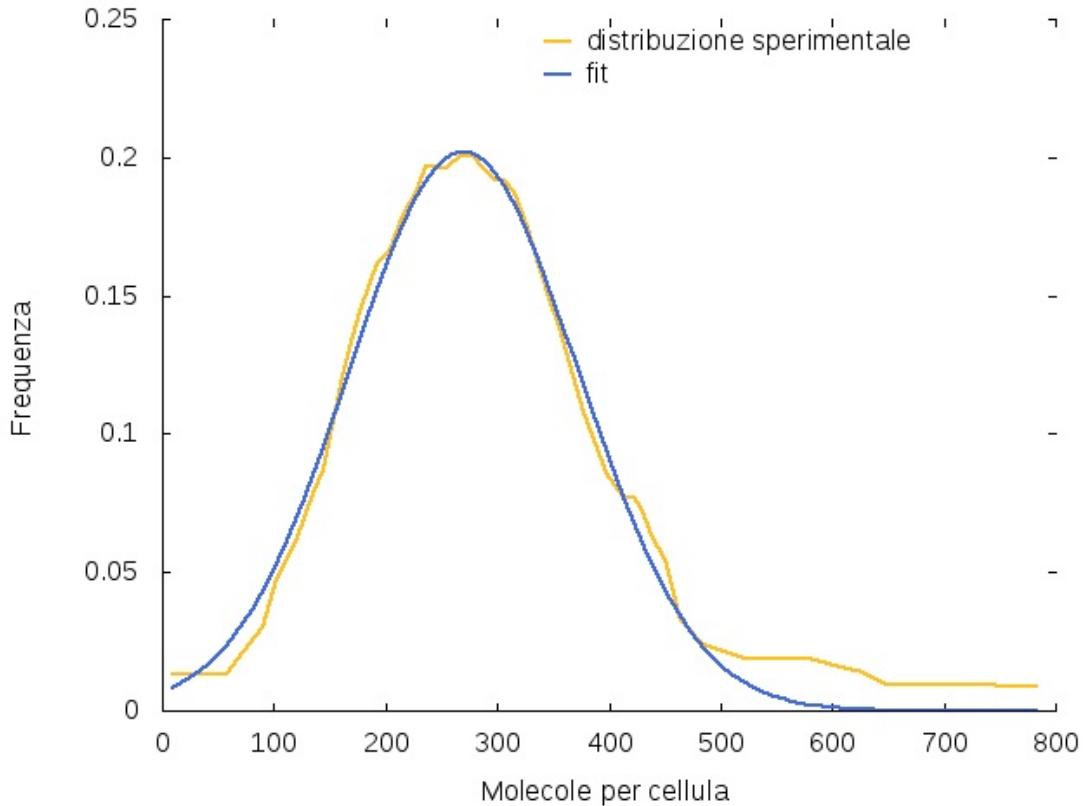


Figura 16: **Distribuzione di probabilità citoplasma** Curva dati sperimentali in giallo e il relativo fit in azzurro. Sulle ascisse è riportata la quantità di mRNA maturo per cellula misurata allo stato stazionario, sulle ordinate la frequenza.

distribuzione dei dati sperimentali non è gaussiana e che si avvicina alla (5.1).

In ultima analisi si vuole dimostrare, per il momento solo numericamente, come la conoscenza dei primi momenti sia sufficiente per determinare le espressioni dei rate poiché i momenti di ordine superiore sono affetti da errori relativi troppo alti e quindi perdono di significato.

Sono stati considerati gli $N=1000$ istogrammi generati con il metodo *bootstrap* secondo la distribuzione sperimentale in figura 6 assunta come distribuzione vera, per ognuno di essi sono stati calcolati, mediante l'equazione (4.1), i primi 5 momenti della distribuzione di probabilità.

È stato calcolato il valore medio per i primi cinque momenti e la relativa incertezza data dalla deviazione standard della media.

In tabella 13 sono riportati i valori dei primi cinque momenti con i rispettivi errori relativi percentuali.

4. Distribuzione di probabilità del citoplasma

$\langle m \rangle$	328	2,3 %
$\langle m^2 \rangle$	117812	3,765%
$\langle m^3 \rangle$	49423715,76	5,77%
$\langle m^4 \rangle$	23832105354	8,455%
$\langle m^5 \rangle$	1.29E+013	10,5%

Tabella 13: Primi cinque momenti della distribuzione di mRNA nel citoplasma con rispettivi errori relativi percentuali.

Gli errori relativi percentuali aumentano all'aumentare dell'ordine dei momenti, non si nota però un brusco aumento tale da poter trascurare i momenti di ordine superiore. Per riprodurre fedelmente la distribuzione di probabilità sperimentale sarà quindi necessario considerare anche i momenti di ordine maggiore.

Conclusione

In questo lavoro è stato elaborato un modello teorico per la trascrizione genica che a differenza dei modelli presenti in letteratura [3] permette di ricavare informazioni sul fenomeno della trascrizione non solo a livello del citoplasma ma anche a livello del nucleo.

Dal modello, sviluppato utilizzando la teoria delle master equation, infatti è possibile ricavare le espressioni per il rate g di produzione dell'mRNA nel nucleo, per i rate w_+ , w_- che regolano l'attività del promotore e per il rate r che indica il tasso di esportazione dell'mRNA dal nucleo al citoplasma, in funzione dei soli primi due momenti della distribuzione di probabilità di nucleo e citoplasma che verranno determinati a partire dalle distribuzioni sperimentali.

Con queste equazioni sarà possibile determinare direttamente i rate e descrivere quindi il processo di trascrizione, facendo riferimento esclusivamente alle distribuzioni sperimentali della quantità di mRNA nel nucleo e nel citoplasma allo stato stazionario.

Ci si è occupati inoltre di studiare l'andamento dei rate al variare del tempo del ciclo cellulare; dall'analisi svolta è emerso che i rate g e r , variano linearmente al variare del tempo del ciclo cellulare: questo porterebbe a ipotizzare che durante il ciclo cellulare viene messo in atto un meccanismo di risposta all'aumento della produzione di mRNA che corrisponde a un aumento della velocità del processo trascrizionale.

Il meccanismo di regolazione, che è stato supposto, avviene soprattutto a livello del nucleo dal momento che il rate g ha un andamento lineare più marcato rispetto al rate r .

Per quanto riguarda i rate w_+ , w_- , non si osserva una andamento lineare: durante tutto il ciclo cellulare il promotore può essere modellizzato con il processo Random Telegraph.

A Processo di Markov

Un processo di Markov è un processo stocastico con probabilità di transizione a uno stato del sistema dipendente solo dallo stato del sistema immediatamente precedente [2].

Per ogni insieme di n successivi tempi $t_1 < t_2 < \dots < t_n$ vale la proprietà di Markov:

$$P_{1|n-1}(y_n, t_n | y_1, t_1; \dots; y_{n-1}, t_{n-1}) = P_{1|1}(y_n, t_n | y_{n-1}, t_{n-1}) \quad (\text{A.1})$$

cioè la densità di probabilità condizionata al tempo t_n dato il valore y_{n-1} al tempo t_{n-1} , è univocamente determinata e non dipende dalla conoscenza dei valori ai tempi precedenti.

Prendendo per esempio $t_1 < t_2 < t_3$ la probabilità si può esprimere nel seguente modo:

$$\begin{aligned} P_3(y_1, t_1; y_2, t_2; y_3, t_3) &= P_2(y_1, t_1; y_2, t_2) P_{1|2}(y_3, t_3 | y_1, t_1; y_2, t_2) \\ &= P_1(y_1, t_1) P_{1|1}(y_2, t_2 | y_1, t_1) P_{1|1}(y_3, t_3 | y_2, t_2) \end{aligned}$$

Un processo di Markov è univocamente determinato dalla conoscenza di $P_1(y_1, t_1)$ e $P_{1|1}(y_2, t_2 | y_1, t_1)$.

Integrando l'equazione (A.2) su y_2 per tempi $t_1 < t_2 < t_3$:

$$P_2(y_1, t_1; y_3, t_3) = P_1(y_1, t_1) \int P_{1|1}(y_2, t_2 | y_1, t_1) P_{1|1}(y_3, t_3 | y_2, t_2) dy_2 \quad (\text{A.2})$$

e dividendo entrambi i membri per $P_1(y_1, t_1)$, si ottiene l'equazione di *Chapman-Kolmogorov*:

$$P_{1|1}(y_3, t_3 | y_1, t_1) = \int P_{1|1}(y_3, t_3 | y_2, t_2) P_{1|1}(y_2, t_2 | y_1, t_1) dy_2 \quad (\text{A.3})$$

Si consideri un processo di Markov stazionario, la probabilità di transizione $P_{1|1}$ è indipendente rispetto ai due tempi e dipende solo dall'intervallo di tempo. Introducendo la seguente notazione:

$$P_{1|1}(y_2, t_2 | y_1, t_1) = T_\tau(y_2 | y_1) \quad \text{con } \tau = \int t_2 - t_1$$

l'equazione di *Chapman-Kolmogorov* diventa (per $\tau' > 0$)

$$T_{\tau+\tau'}(y_3 | y_1) = \int T_{\tau'}(y_3 | y_2) T_\tau(y_2 | y_1) dy_2 \quad (\text{A.4})$$

Appendice

A.1 Ricavazione Master Equation

Si consideri un processo di Markov omogeneo, così che si possa scrivere la T_τ per esprimere la probabilità di transizione, si può dimostrare che $T_\tau(y_2|y_1)$ per tempi τ' piccoli presenta due contributi:

$$T_{\tau'}(y_2|y_1) = (1 - a_0\tau')\delta(y_2 - y_1) + \tau'W(y_2|y_1) + o(\tau') \quad (\text{A.5})$$

il primo termine indica la probabilità che non avvengano transizioni durante τ' , mentre il secondo termine indica la probabilità di transizione dallo stato y_1 a y_2 con $W(y_2|y_1)$ rate di transizione definito positivo.

a_0 è definito nel seguente modo:

$$a_0(y_1) = \int W(y_2|y_1)dy_2 \quad (\text{A.6})$$

Inserendo l'espressione appena trovata nell'equazione (A.4), si ottiene:

$$T_{\tau+\tau'}(y_3|y_1) = [1 - a_0(y_3)\tau'] T_\tau(y_3|y_1) + \tau' \int W(y_3|y_2)T_\tau(y_2|y_1)dy_2 \quad (\text{A.7})$$

Sostituendo il termine a_0 trovato precedentemente e dividendo per τ' , dopo aver svolto il limite per $\tau' \rightarrow 0$ si ottiene:

$$\frac{\partial T_\tau(y_3|y_1)}{\partial \tau} = \int \left\{ W(y_3|y_2)T_\tau(y_2|y_1) - W(y_2|y_3)T_\tau(y_3|y_1) \right\} dy_2 \quad (\text{A.8})$$

La forma differenziale dell'equazione di Chapman-Kolmogorov chiamata *master equation*, solitamente viene scritta in una forma più intuitiva nel seguente modo:

$$\frac{\partial P(y,t)}{\partial \tau} = \int \left\{ W(y|y')P(y',t) - W(y'|y)P(y,t) \right\} dy' \quad (\text{A.9})$$

Si prenda un tempo t_1 e un valore y_1 e si consideri la soluzione che è determinata per $t \geq t_1$ dalla condizione iniziale $P(y,t_1) = \delta(t-t_1)$. La soluzione è la probabilità di transizione $T_{t-t_1}(y|y_1)$ del processo di markov $\forall t_1, y_1$. Se il range di Y è un insieme discreto di stati indicizzati con n , l'equazione si riduce a:

$$\frac{dP_n(t)}{dt} = \sum_{n'} \left\{ W_{nn'}p_{n'}(t) - W_{n'n}p_n(t) \right\} \quad (\text{A.10})$$

Il primo termine rappresenta il guadagno di stati n dovuto alla transizione da altri stati n' , il secondo termine rappresenta la perdita dovuta al passaggio da stati n ad altri stati.

Appendice

A.2 Processo one-step

È un processo di Markov a tempo continuo, il cui spazio degli stati è rappresentato dalla variabile $n = (n_1, n_2, \dots, n_L)$ che indica le copie delle L variabili che sto considerando. Si consideri il caso di una sola specie.

Le uniche transizioni permesse per questo processo sono quelle tra siti adiacenti, per cui la matrice di transizione può essere scritta come:

$$W_{nn'} = r_{n'}\delta_{n,n'-1} + g_{n'}\delta_{n,n'+1} \quad \text{con } (n \neq n')$$

Gli elementi sulla diagonale hanno la forma $W_{nn'} = -(r_n + g_n)$. Gli unici elementi che contribuiscono alla matrice $W_{nn'}$ sono:

$$W_{n+1,n'}(t) = g_n \quad \text{con } g_n \text{ rate di nascita costante, } g_n = g$$

$$W_{n-1,n'}(t) = r_n \quad \text{con } r_n \text{ rate di morte proporzionale a } n, r_n = nr$$

Sostituendo gli elementi di matrice appena trovati nell'equazione (A.10) si ottiene la master equation

$$\langle p_n \rangle = -gp_n - np_n + gp_{n-1} + (n+1)p_{n+1} \quad (\text{A.11})$$

B Processo Random-Telegraph

Questo processo è un processo stocastico a tempo continuo, il cui spazio degli stati è definito da due valori a, b . Il processo può essere descritto dalle seguenti master equation:

$$\begin{aligned} \partial_t P(a, t|x, t_0) &= -\lambda P(a, t|x, t_0) + \mu P(b, t|x, t_0) \\ \partial_t P(b, t|x, t_0) &= +\lambda P(a, t|x, t_0) - \mu P(b, t|x, t_0) \end{aligned} \quad (\text{B.1})$$

Il promotore alterna casualmente lo stato in cui è attivo e quello in cui è spento, perciò il suo comportamento può essere modellizzato mediante questo processo; la variabile π assuma il valore 0 quando il promoter è spento e il valore 1 quando il promoter è attivo.

C Funzione generatrice

In questa sezione è riportato dettagliatamente il calcolo per passare dalla master equation all'equazione differenziale per la funzione generatrice svolto nel primo

Appendice

capitolo.

Per lo stato attivo si ottiene:

$$\begin{aligned}
0 &= g \sum_{m'm} p_{m',m}^+ y^{m'+1} x^m - g \sum_{m'm} p_{m',m}^+ y^{m'} x^m + r \sum_{m'm} (m') p_{m',m}^+ y^{m'-1} x^{m+1} \\
&\quad - rm' p_{m',m}^+ y^{m'} x^m + \gamma \sum_{m'm} (m) p_{m',m+1}^+ y^{m'} x^{m-1} - \gamma \sum_{m'm} mp_{m',m}^+ y^{m'} x^m \\
&\quad + w_+ \sum_{m'm} p_{m',m}^- y^{m'} x^m - w_- \sum_{m'm} p_{m',m}^+ y^{m'} x^m \\
0 &= gy \sum_{m'm} p_{m',m}^+ y^{m'} x^m - g \sum_{m'm} p_{m',m}^+ y^{m'} x^m + rx \sum_{m'm} p_{m',m}^+ x^m \partial_y y^{m'} \\
&\quad - ry \sum_{m'm} p_{m',m}^+ x^m \partial_y y^{m'} + \gamma \sum_{m'm} p_{m',m}^+ y^{m'} \partial_x x^m - \gamma x \sum_{m'm} p_{m',m}^+ y^{m'} \partial_x x^m \\
&\quad + w_+ \sum_{m'm} p_{m',m}^- y^{m'} x^m - w_- \sum_{m'm} p_{m',m}^+ y^{m'} x^m \\
0 &= gyG_+, -gG_+ + rx\partial_y G_+ - ry\partial_y G_+ + \gamma\partial_x G_+ - \gamma x\partial_x G_+ + w_+G_- - w_-G_+
\end{aligned}$$

Per lo stato inattivo si ottiene:

$$\begin{aligned}
0 &= r \sum_{m'm} (m'+1) p_{m'+1,m-1}^- y^{m'} x^m - rm' p_{m',m}^- y^{m'} x^m + \gamma \sum_{m'm} (m+1) p_{m',m+1}^- y^{m'} x^m \\
&\quad - \gamma \sum_{m'm} mp_{m',m}^- y^{m'} x^m - w_+ \sum_{m'm} p_{m',m}^- y^{m'} x^m + w_- \sum_{m'm} p_{m',m}^+ y^{m'} x^m \\
0 &= r \sum_{m'm} (m') p_{m',m}^- y^{m'-1} x^{m+1} - rm' p_{m',m}^- y^{m'} x^m + \gamma \sum_{m'm} (m) p_{m',m+1}^- y^{m'} x^{m-1} \\
&\quad - \gamma \sum_{m'm} mp_{m',m}^- y^{m'} x^m - w_+ \sum_{m'm} p_{m',m}^- y^{m'} x^m + w_- \sum_{m'm} p_{m',m}^+ y^{m'} x^m \\
0 &= rx \sum_{m'm} p_{m',m}^- x^m \partial_y y^{m'} - ry \sum_{m'm} p_{m',m}^- x^m \partial_y y^{m'} + \gamma \sum_{m'm} p_{m',m}^- y^{m'} \partial_x x^m \\
&\quad - \gamma x \sum_{m'm} p_{m',m}^- y^{m'} \partial_x x^m - w_+ \sum_{m'm} p_{m',m}^- y^{m'} x^m + w_- \sum_{m'm} p_{m',m}^+ y^{m'} x^m \\
0 &= rx\partial_y G_- - ry\partial_y G_- + \gamma\partial_x G_- - \gamma x\partial_x G_- - w_+G_- + w_-G_+
\end{aligned}$$

D Bootstrap

Il bootstrap è una tecnica statistica di ricampionamento con reimmissione utile per calcolare valori medi, varianza e p-values quando non si ha a disposizione una singola misura della distribuzione di probabilità (quando non si conosce la distribuzione della statistica di interesse). Si consideri un campione effettivamente osservato di numerosità pari ad n , diciamo $x = (x_1, \dots, x_n)$. Da x si ricampionano altri m campioni di numerosità costante pari ad n , diciamo x_1^*, \dots, x_m^* . In ciascuna

Appendice

estrazione bootstrap, i dati provenienti dal primo elemento del campione ,cioè x_1 ,possono essere estratti piú di una volta e ciascun dato ha probabilità pari a $1/n$ di essere estratto, mentre se voglio generare dati che seguano una certa distribuzione, allora ogni dato sarà estratto con probabilità pari alla probabilità del relativo bin. Sia T lo stimatore di θ che ci interessa studiare, diciamo $T(x) = \theta$.Si calcoli tale quantità per ogni campione bootstrap, $T_{x_1^*}, \dots T_{x_m^*}$.In questo modo si hanno a disposizione m stime di θ , dalle quali è possibile calcolare la media bootstrap, la varianza bootstrap,il coefficiente di correlazione lineare r ; Partendo quindi da queste quantità stimate è possibile calcolare il p-values di una certa ipotesi nulla.

E Espressioni analitiche rate

A partire dalle equazioni riportate sotto, si determinano le espressioni analitiche dei rate in funzione dei momenti della distribuzione di probabilità.

$$\begin{aligned}
\langle m \rangle &= \frac{g}{\gamma} \langle \pi \rangle \\
\langle m' \rangle &= \frac{g}{r} \langle \pi \rangle \\
\langle m'^2 \rangle &= \frac{1}{2r}(g \langle \pi \rangle + r \langle m' \rangle + 2g \langle m' \rangle_+) \\
\langle m'm \rangle &= \frac{g \langle m \rangle_+ + r \langle m'^2 \rangle - r \langle m' \rangle}{r + \gamma} \\
\langle m^2 \rangle &= \frac{1}{2\gamma}(r \langle m' \rangle + 2r \langle mm' \rangle + \gamma \langle m \rangle) \\
\langle \pi \rangle &= \frac{w_+}{w_+ + w_-}
\end{aligned} \tag{E.1}$$

Si ricavi dalla quinta equazione delle (E.1) il valore medio $\langle mm' \rangle$, il cui valore numerico è noto poichè si conoscono i momenti centrali della distribuzione di mRNA nel citoplasma e nel nucleo.

$$\langle mm' \rangle = \frac{\langle m' \rangle (\langle m^2 \rangle - \langle m \rangle)}{\langle m \rangle} \tag{E.2}$$

Appendice

Si determini il valor medio $\langle m' \rangle_+$ presente nella terza equazione della E.1 a partire dell'espressione della master equation per lo stato attivo (3.1):

$$\begin{aligned}
\dot{\langle m' \rangle}_+ &= g \sum_{m'm} m' p_{m'-1,m}^+ - g \sum_{m'm} m' p_{m',m}^+ + r \sum_{m'm} m'(m'+1) p_{m'+1,m-1}^+ - r \sum_{m'm} m'^2 p_{m',m}^+ \\
&\quad + \gamma \sum_{m'm} m'(m+1) p_{m',m+1}^\pi - \gamma \sum_{m'm} mm' p_{m',m}^+ - w_- \sum_{m'm} m' p_{m',m}^+ + w_+ \sum_{m'm} m' p_{m',m}^- \\
&= g \sum_{m'm} (m'+1) p_{m',m}^+ - g \sum_{m'm} m' p_{m',m}^+ + r \sum_{m'm} m'(m'-1) p_{m',m}^+ - r \sum_{m'm} m'^2 p_{m',m}^+ \\
&\quad + \gamma \sum_{m'm} m'(m) p_{m',m}^\pi - \gamma \sum_{m'm} mm' p_{m',m}^+ - w_- \sum_{m'm} m' p_{m',m}^+ + w_+ \sum_{m'm} m' p_{m',m}^- \\
&= g \langle \pi \rangle - r \langle m'_+ \rangle + w_+ \langle m' \rangle_- - w_- \langle m' \rangle_+
\end{aligned} \tag{E.3}$$

Analogamente si determini il valor medio $\langle m' \rangle_-$ a partire dalla master equation per lo stato inattivo (3.2):

$$\dot{\langle m' \rangle}_- = \frac{w_- \langle m' \rangle_+}{(r + w_+)} \tag{E.4}$$

Si sostituisce l'espressione appena ricavata nell'equazione (E.3) e si determini il valor medio $\langle m' \rangle_+$.

$$\langle m' \rangle_+ = \frac{g \langle \pi \rangle (r + w_+)}{r(r + w_+ + w_-)} \tag{E.5}$$

Si determini il valor medio $\langle m \rangle_+$ e $\langle m \rangle_-$ a partire rispettivamente dalla master equation per lo stato attivo (3.1) e per lo stato inattivo (3.2):

$$\begin{aligned}
\dot{\langle m \rangle}_+ &= r \langle m' \rangle_+ - \gamma \langle m \rangle_+ + w_+ \langle m \rangle_- - w_- \langle m \rangle_+ \\
\dot{\langle m \rangle}_- &= r \langle m' \rangle_- - \gamma \langle m \rangle_- + w_- \langle m \rangle_+ - w_+ \langle m \rangle_-
\end{aligned}$$

Allo stato stazionario si ottiene:

$$\langle m \rangle_- = \frac{r \langle m' \rangle_- + w_- \langle m' \rangle_+}{(\gamma + w_+)}$$

Si risolva infine il seguente sistema:

$$\begin{cases}
\langle m \rangle_- - \frac{r \langle m' \rangle_- + w_- \langle m' \rangle_+}{(\gamma + w_+)} = 0 \\
r \langle m' \rangle_- - \gamma \langle m \rangle_- + w_- \langle m \rangle_+ - w_+ \langle m \rangle_- = 0 \\
g \langle m \rangle_+ + r \langle m'^2 \rangle - r \langle m' \rangle - \gamma \langle mm' \rangle - r \langle mm' \rangle = 0 \\
\langle m'^2 \rangle - \frac{1}{2r}(g \langle \pi \rangle + r \langle m' \rangle + 2g \langle m' \rangle_+) = 0
\end{cases}$$

con incognite $w_+, w_-, \langle m \rangle_+$ e $\langle m \rangle_-$.

$$\langle m \rangle_+ = \frac{N}{D} \tag{E.6}$$

Appendice

$$N = 4\gamma r(\langle m \rangle \langle m'^2 \rangle r - \langle m' \rangle (\langle m^2 \rangle (\gamma + r) - \gamma \langle m \rangle))(\gamma \langle m \rangle^3 - \langle m' \rangle (\langle m^2 \rangle (\gamma + r) - \gamma \langle m \rangle) + \langle m \rangle \langle m'^2 \rangle r)$$

$$\begin{aligned} D = & 4\gamma^4 \langle m \rangle^2 \langle m' \rangle (\langle m^2 \rangle - \langle m \rangle) + 2\gamma \langle m \rangle r^3 (\langle m' \rangle (\langle m \rangle^2 \\ & + \langle m \rangle + \langle m' \rangle - \langle m^2 \rangle) - \langle m'^2 \rangle (2 \langle m \rangle^2 + \langle m \rangle + 3 \langle m' \rangle) + 2 \langle m'^2 \rangle^2) + \\ & 2\gamma^3 \langle m \rangle r (\langle m \rangle^3 - 2 \langle m \rangle^2 \langle m'^2 \rangle + \langle m \rangle \langle m' \rangle (2 \langle m^2 \rangle - 1) \\ & + \langle m' \rangle \langle m^2 \rangle) + \gamma^2 r^2 (2 \langle m \rangle^4 + \langle m \rangle^3 (2 \langle m' \rangle - 4 \langle m'^2 \rangle + 1) \\ & + 2 \langle m \rangle^2 (\langle m' \rangle - \langle m'^2 \rangle) - 2 \langle m \rangle \langle m' \rangle (\langle m' \rangle - 2 \langle m'^2 \rangle) + 2 \langle m' \rangle \langle m^2 \rangle \\ & (\langle m' \rangle - 2 \langle m'^2 \rangle)) + \langle m' \rangle r^4 (\langle m \rangle - 2 \langle m^2 \rangle) (\langle m' \rangle - 2 \langle m'^2 \rangle) \end{aligned}$$

$$\langle m \rangle_- = \frac{N}{D} \tag{E.7}$$

$$\tag{E.8}$$

$$\begin{aligned} N = & ((\langle m \rangle^2 r\gamma - 2 \langle m' \rangle \langle m^2 \rangle r(r + \gamma) + 2 \langle m \rangle^3 \gamma(r + \gamma) \\ & + \langle m \rangle \langle m' \rangle r(r + 2\gamma))(\langle m \rangle^2 r\gamma + 2 \langle m' \rangle \langle m^2 \rangle \gamma(r + \gamma) \\ & + \langle m \rangle (-2 \langle m'^2 \rangle r(r + \gamma) + \langle m' \rangle (r^2 - 2\gamma^2)))) \end{aligned}$$

$$\begin{aligned} D = & (\langle m' \rangle (\langle m \rangle - 2 \langle m^2 \rangle)(\langle m' \rangle - 2 \langle m'^2 \rangle)r^4 + 2 \langle m \rangle (\langle m' \rangle (\langle m \rangle + \\ & \langle m \rangle^2 + \langle m' \rangle - \langle m^2 \rangle) - (\langle m \rangle + 2 \langle m \rangle^2 + 3 \langle m' \rangle) \langle m'^2 \rangle + 2 \langle m'^2 \rangle^2)r^3\gamma \\ & + (2 \langle m \rangle^4 + \langle m \rangle^3 (1 + 2 \langle m' \rangle - 4 \langle m'^2 \rangle) - 2 \langle m \rangle \langle m' \rangle (\langle m' \rangle - 2 \langle m'^2 \rangle) \\ & + 2 \langle m' \rangle \langle m^2 \rangle (\langle m' \rangle - 2 \langle m'^2 \rangle) + 2 \langle m \rangle^2 (\langle m' \rangle - \langle m'^2 \rangle))r^2\gamma^2 \\ & + 2 \langle m \rangle (\langle m \rangle^3 + \langle m' \rangle \langle m^2 \rangle + \langle m \rangle \langle m' \rangle (-1 + 2 \langle m^2 \rangle) \\ & - 2 \langle m \rangle^2 \langle m'^2 \rangle)r\gamma^3 + 4 \langle m \rangle^2 \langle m' \rangle (-\langle m \rangle + \langle m^2 \rangle)\gamma^4) \end{aligned}$$

$$w_+ = \frac{N}{D} \tag{E.9}$$

$$\tag{E.10}$$

$$\begin{aligned} N = & -((2 \langle m \rangle^2 r\gamma^2 (\langle m \rangle^2 r\gamma + 2 \langle m' \rangle \langle m^2 \rangle \gamma(r + \gamma) \\ & + \langle m \rangle (-2 \langle m'^2 \rangle r(r + \gamma) + \langle m' \rangle (r^2 - 2\gamma^2)))) \end{aligned}$$

Appendice

$$\begin{aligned}
D = & (\langle m' \rangle (\langle m \rangle - 2 \langle m^2 \rangle) (\langle m' \rangle - 2 \langle m'^2 \rangle) r^4 + 2 \langle m \rangle (\langle m' \rangle (\langle m \rangle \\
& + \langle m \rangle^2 + \langle m' \rangle - \langle m^2 \rangle) - (\langle m \rangle + 2 \langle m \rangle^2 + 3 \langle m' \rangle) \langle m'^2 \rangle + 2 \langle m'^2 \rangle^2) \\
& r^3 \gamma + (2 \langle m \rangle^4 + \langle m \rangle^3 (1 + 2 \langle m' \rangle - 4 \langle m'^2 \rangle) - 2 \langle m \rangle \langle m' \rangle (\langle m' \rangle \\
& - 2 \langle m'^2 \rangle) + 2 \langle m' \rangle \langle m^2 \rangle (\langle m' \rangle - 2 \langle m'^2 \rangle) + 2 \langle m \rangle^2 (\langle m' \rangle - \langle m'^2 \rangle)) \\
& r^2 \gamma^2 + 2 \langle m \rangle (\langle m \rangle^3 + \langle m' \rangle \langle m^2 \rangle + \langle m \rangle \langle m' \rangle (-1 + 2 \langle m^2 \rangle) \\
& - 2 \langle m \rangle^2 \langle m'^2 \rangle) r \gamma^3 + 4 \langle m \rangle^2 \langle m' \rangle (-\langle m \rangle + \langle m^2 \rangle) \gamma^4))
\end{aligned}$$

$$w_- = \frac{N}{D} \quad (E.11)$$

$$(E.12)$$

$$\begin{aligned}
N = & ((\langle m' \rangle (\langle m \rangle - 2 \langle m^2 \rangle) r^2 + \langle m \rangle (\langle m \rangle + 2 \langle m \rangle^2 + 2 \langle m' \rangle - 2 \langle m'^2 \rangle) \\
& r \gamma + 2 \langle m' \rangle (-\langle m \rangle + \langle m^2 \rangle) \gamma^2) (\langle m' \rangle r^2 - 2 \langle m'^2 \rangle r^2 + \\
& \langle m \rangle \gamma (r + 2 \langle m \rangle \gamma)) (\langle m \rangle^2 r \gamma + 2 \langle m' \rangle \langle m^2 \rangle \gamma (r + \gamma) \\
& + \langle m \rangle (-2 \langle m'^2 \rangle r (r + \gamma) + \langle m' \rangle (r^2 - 2 \gamma^2))))
\end{aligned}$$

$$\begin{aligned}
D = & (2(\langle m' \rangle (\langle m \rangle - 2 \langle m^2 \rangle) (\langle m' \rangle - 2 \langle m'^2 \rangle) r^4 + 2 \langle m \rangle (\langle m' \rangle (\langle m \rangle \\
& + \langle m \rangle^2 + \langle m' \rangle - \langle m^2 \rangle) - (\langle m \rangle + 2 \langle m \rangle^2 + 3 \langle m' \rangle) \langle m'^2 \rangle \\
& + 2 \langle m'^2 \rangle^2) r^3 \gamma + (2 \langle m \rangle^4 + \langle m \rangle^3 (1 + 2 \langle m' \rangle - 4 \langle m'^2 \rangle) \\
& - 2 \langle m \rangle \langle m' \rangle (\langle m' \rangle - 2 \langle m'^2 \rangle) + 2 \langle m' \rangle \langle m^2 \rangle \\
& (\langle m' \rangle - 2 \langle m'^2 \rangle) + 2 \langle m \rangle^2 (\langle m' \rangle - \langle m'^2 \rangle)) r^2 \gamma^2 + 2 \langle m \rangle (\langle m \rangle^3 \\
& + \langle m' \rangle \langle m^2 \rangle + \langle m \rangle \langle m' \rangle (-1 + 2 \langle m^2 \rangle) - 2 \langle m \rangle^2 \langle m'^2 \rangle) r \gamma^3 + 4 \langle m \rangle^2 \\
& \langle m' \rangle (-\langle m \rangle + \langle m^2 \rangle) \gamma^4) (\langle m \rangle \langle m'^2 \rangle r + \langle m \rangle^3 \gamma - \langle m' \rangle \\
& (-\langle m \rangle \gamma + \langle m^2 \rangle (r + \gamma)))))
\end{aligned}$$

Dalla prima e della seconda delle E.1 si determinino i rate r e g :

$$\begin{aligned}
r &= \frac{\langle m \rangle}{\langle m' \rangle} \gamma \\
g &= \frac{\gamma \langle m \rangle}{\langle \pi \rangle}
\end{aligned} \quad (E.13)$$

Per determinare l'espressione analitica per il rate g si sostituisca nella seconda equazione (E.13) l'espressioni per i rate w_+ e w_- appena trovate.

Riferimenti bibliografici

- [1] S. O. Skinner, H. Xu, S. Nagarkar-Jaiswal, P. R Freire, T. P Zwaka and I. Golding. Single-cell analysis of transcription kinetics across the cell cycle. *eLife*, 2016.
- [2] N. G. Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.
- [3] A. M. Walczak, A. Mugler, and C. H. Wiggins. Analytic methods for modeling stochastic regulatory networks. *Computational Modeling of Signaling Networks*, Volume 880 *Methods in Molecular Biology*, p. 273-322, 2012.
- [4] H. Xu, S. O. Skinner, A. M. Sokac and I. Golding. Stochastic Kinetics of Nascent RNA. *Physical Review Letters*, 117 (12), 128101, 2016
- [5] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas and S. Tyagi. Stochastic mRNA Synthesis in Mammalian Cells. *PLOS Biology*, 4(10), 2006.
- [6] R. Phillips, J. Kondev, J. Theriot and H. G. Garcia. Physical Biology of the Cells. *Garland Science*, seconda edizione, , p. 253-256.
- [7] E.T.Jaynes. Information Theory and statistical Mechanics. *PHISICAL REVIEW*, volume 106,number 4.California,1957.
- [8] B.Alberts, D.Bray ,J. Lewis ,M. Raff, K. Roberts and J. D. Watson. Biologia Molecolare della Cellula. *Zanichelli*, Bologna ,1984.