

Relatore:

Prof. Ginelli Francesco

Laureando:

Arrigoni Stefano

Totally Asymmetric Simple Exclusion Process: un modello per lo studio della traduzione del mRNA

Nell'elaborato di natura compilativa si tratta tramite l'Asymmetric Simple Exclusion Process (ASEP) o Totally Asymmetric Simple Exclusion Process (TASEP) la sintesi del polipeptide, sequenza di amminoacidi che formeranno in seguito ad altri processi chimici una proteina. Il modello TASEP permette lo studio della cinetica della traduzione del mRNA (RNA messaggero) definendo il ribosoma (complesso macromolecolare formato da 4 molecole di rRNA e 80 proteine) come una particella che si muove lungo un reticolo unidimensionale, il filamento di mRNA stesso.

Il processo di traduzione è un processo sequenziale nel quale un ribosoma scorre lungo il filamento di mRNA. Ad ogni "tappa" si verifica l'accoppiamento di uno specifico tRNA (RNA di trasferimento) con il codone (unione di 3 basi azotate che codifica per uno specifico amminoacido) presente sul filamento di RNA messaggero fino all'abbandono del tRNA stesso dopo il rilascio dell'amminoacido che si lega al polipeptide nel processo di elongazione. Il modello TASEP descrive lo spostamento di particelle con una dinamica stocastica in una sola direzione lungo un reticolo unidimensionale, riprendendo appunto la sequenzialità del processo biologico. Il vincolo di esclusione inoltre riprende l'impossibilità di più ribosomi, impegnati contemporaneamente nel processo di traduzione, di sovrapporsi sullo stesso codone.

La parte introduttiva approfondisce le proprietà biologiche e descrive i modelli originali introdotti negli anni '68 e '69 che, attraverso una semplice approssimazione di campo medio trascurando la correlazione tra i diversi siti, permettono di descrivere analiticamente le differenze di fase che possono verificarsi nel sistema (fase ad alta densità, fase a bassa densità, fase di massima corrente), i punti critici che definiscono le transizioni di fase, la densità media (ρ) e locale (ρ_i) delle particelle tra i vari siti e la corrente (J) di spostamento delle particelle. Il sistema è trattato in due diverse condizioni al contorno: *condizioni al contorno periodiche* in cui il filamento di mRNA è rappresentato come un anello e permette una trattazione più semplice, ma meno verosimile; *condizioni al contorno aperte* che rappresenta in modo più corretto l'ambiente citoplasmatico in cui i tRNA diffondono. La dinamica nel modello TASEP è governata da un processo stocastico in cui viene scelto in maniera randomica un sito k , se il sito è pieno (ovvero ha una particella al suo interno, o, nel processo biologico, un ribosoma si trova sul codone in posizione k) e il sito $k+1$ è vuoto allora la particella salterà dal sito k al sito $k+1$. La dinamica è quindi riassunta nel modello TASEP definendo un rate di salto tra un sito e l'altro che, a differenza del modello inomogeneo (trattato nell'ultimo capitolo), è assunto costante lungo tutto il reticolo. Inoltre nel caso di condizioni al contorno aperte viene introdotto un rate di immissione della particella nel reticolo (α) e un rate di espulsione della particella dal reticolo (β).

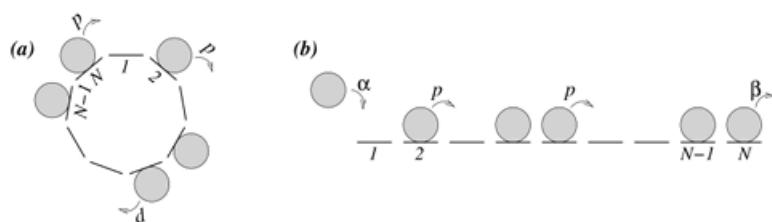


Figura 1: Nella figura è possibile vedere una rappresentazione schematica del modello TASEP. (a) coincide con la condizione di periodic boundary, (b) rappresenta la condizione di open boundary

Nel terzo capitolo si mette in evidenza l'importanza del processo di accoppiamento tra tRNA e codone, poiché i dati sperimentali indicano che corrisponde al processo più lento limitante la cinetica. Questa proprietà viene trattata aggiungendo un grado di libertà interno (o stato interno). La descrizione del processo biologico viene quindi riassunta in due stati principali: lo stato di accoppiamento (primo stato) e lo stato di traslocazione (secondo stato). I calcoli vengono semplificati assumendo l'ambiente omogeneo, ovvero assumendo la presenza omogenea dei tRNA nell'ambiente citoplasmatico, infatti in questo modo è possibile trascurare le fluttuazioni di densità. Inoltre, attraverso l'approssimazione di campo medio viene trascurata la correlazione tra i diversi siti.

Il modello a due stati ha messo in evidenza che i punti critici in cui avvengono le transizioni di fase sono diversi da quelli predetti dal modello precedente. In particolare, la fase di corrente massima risulta ora preponderante nel diagramma di fase, con i punti critici e le variazioni di fase che dipendono dal rapporto tra k (rate che

definisce la transizione da primo a secondo stato, processo di accoppiamento) e γ (rate che rappresenta la traslocazione del ribosoma da un sito ad un altro del reticolo).

A valori di k crescenti la fase di massima corrente (MC phase) si riduce fino alla condizione $\frac{k}{\gamma} \rightarrow \infty$ per cui si ottengono i valori descritti dal TASEP originale. La condizione limite coincide col rappresentare il passaggio da primo stato a secondo istantaneo.

Dai dati sperimentali emerge che il valore reale di γ è dell'ordine di 35 s^{-1} (costante per ogni codone) e il rapporto $\frac{k}{\gamma}$ si trova nel range [0.05-3.38] in dipendenza del codone scelto. Tuttavia, per la grande maggioranza dei codoni (87%) si ha, dai dati sperimentali, che $k < 1$. Infatti il risultato principale del modello a due stati è dimostrare che il processo di traduzione giace quasi costantemente in una fase di corrente massima, in cui la sintesi delle proteine è ottimizzata, come è possibile vedere dal diagramma di fase in figura 2.

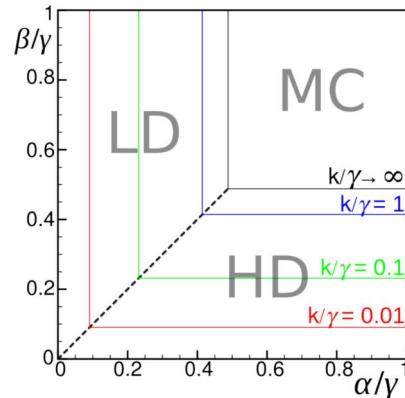


Figura 2: Diagramma di fase in funzione del rapporto $\frac{k}{\gamma}$: maximal current phase (MC); high density phase (HD); low density phase (LD)

Nell'ultimo capitolo si tratta il modello TASEP inomogeneo che si differenzia dai modelli visti in precedenza poiché l'hopping rate (probabilità istantanea di salto) tra un sito e l'altro varia a seconda del sito considerato. Il TASEP inomogeneo descrive la variazione di velocità di scorimento del ribosoma a causa della diversa disponibilità dei tRNA portatori degli aminoacidi associati ai vari codoni. L'omonogenetità dello scorimento può essere rappresentata come una diversa probabilità di salto da un sito ad un altro. È possibile studiare il modello inomogeneo tramite l'approssimazione in serie di potenze (PSA). Il parametro di espansione utilizzato per la PSA è il rate di inizializzazione α che regola l'immissione di nuove particelle sulla catena. Tipicamente infatti quest'ultimo è il processo più lento e dunque α è il rate più piccolo. Si ha quindi:

$$P(C) = \sum_{n=0}^{\infty} c_n \alpha^n \quad \text{dove } P(C) \text{ è la probabilità di trovarsi nella configurazione C del reticolo}$$

L'approssimazione è dominata dai primi termini, dunque studiando i primi K termini dell'espansione e ricavando i primi coefficienti è possibile ottenere il calcolo di densità e corrente con buona approssimazione se K è sufficientemente elevato. Il numero di coefficienti K da calcolare per avere una buona approssimazione come si può intuire è tuttavia enorme. Per questo motivo oltre allo studio del TASEP inomogeneo tramite PSA, nell'ultimo capitolo si approfondisce un pacchetto in codice python, TASEPy, che permette la risoluzione in PSA in modo efficiente e rapido tramite l'utilizzo di: 4 funzioni personalizzate, la definizione della matrice degli hopping rate e la definizione della lunghezza (ℓ) del ribosoma che è visto, nel modello inomogeneo, come un oggetto esteso e la sua posizione è rappresentata dal tracking site (sito centrale nel modello studiato in questo elaborato). TASEPy, inoltre, può essere impiegato per calcoli inversi dei sistemi di traduzione del mRNA. Un primo esempio è il calcolo del rate iniziale α a partire da una data densità sperimentale. A partire dal calcolo dei coefficienti tramite la funzione `psa_compute` è possibile automatizzare il processo utilizzando i tool di ottimizzazione che esplorano i possibili valori del parametro restituendo quello che minimizza la funzione $f(\alpha) = \rho_\alpha - \rho_{exp}$ che misura la differenza di densità corrispondente ad un dato rate α e quella sperimentale. Un secondo impiego di TASEPy permette di calcolare il rapporto tra α e i vari ω_i (gli hopping rate locali) a partire dalla densità locale. Definito $k_i = \frac{\omega_i}{\alpha}$, si possono infatti calcolare i coefficienti tramite la chiamata della funzione `psa_compute` e minimizzando la deviazione quadratica media.

È infine importante notare che l'utilizzo di TASEPy rimane comunque vincolato all'approssimazione in serie di potenze. Per un rate di inizializzazione grande l'approssimazione perde le sue fondamenta restituendo risultati non coerenti.