



UNIVERSITÀ DEGLI STUDI DI MILANO

FACOLTÀ DI SCIENZE E TECNOLOGIE

Corso di Laurea in Fisica

Modellizzazione data-driven della risposta cellulare ad
una transizione dei nutrienti

Tesi di Laurea di:
Valerio Firmano
Matricola: 918239

Relatore:
Marco Cosentino Lagomarsino
Correlatore:
Ludovico Calabrese

Anno Accademico 2020/2021

Contents

I	Introduzione	3
1	Modelli quantitativi di allocazione di risorse per la crescita cellulare	4
1.1	Le "growth laws" condensano la complessa dinamica di crescita cellulare in un modello a settori con pochi parametri	4
1.2	Il modello di regolazione controllata da flussi (FCR) di Erickson si propone di spiegare le transazioni tra stati di crescita cellulare .	7
2	Esperimenti in chip microfluidici permettono di estrarre la dinamica dei settori	9
II	Risultati	12
3	I dati a disposizione permettono di estrarre tutti i parametri rilevanti per le leggi di crescita e le transizioni	12
3.1	Il promotore costitutivo in ter e ribosomale in ori rappresentano bene i rispettivi settori del proteoma.	12
3.2	Il reporter ribosomale P1longori è utilizzabile da solo come proxy 'single strain' per ϕ_R e ϕ_P grazie alla growth law	16
3.3	Lo strain con il miglior proxy per il rate di crescita è P5ter	18
3.3.1	Dati di growth rate di P1longori suddivisi per FOV evidenziano eterogeneità nella crescita	20
3.3.2	Dati di growth rate di P1ori suddivisi per FOV evidenziano eterogeneità nella crescita.	22
3.3.3	Dati di growth rate di P5ter suddivisi per FOV confermano tale strain come miglior proxy	23
4	I nostri dati nel transiente di nutrienti falsificano l'ipotesi di quasi stazionarietà del modello di regolazione controllata dai flussi (FCR)	24
4.1	I dati sono analizzabili con il framework di Erickson e rispettano le leggi di crescita negli stati stazionari	24

4.2	I nostri dati falsificano l'ipotesi di quasi stazionarietà del modello FCR	26
III	Modelli	30
5	Un modello minimale in cui σ e χ_R sono assegnati indipendentemente riproduce qualitativamente i dati nel transiente	30
6	Un modello meccanicistico dell'allocazione dei ribosomi nel transiente devia dall'assunzione di quasi-stazionarietà, ma non riproduce gli overshoot nei dati	32
7	Proposta di un modello con una equazione per la concentrazione di amminoacidi, con σ legato al suo valore	34
IV	Conclusioni	37
A	Metodi	39
A.1	Serie temporali	39
A.2	Estrazione dati Erickson <i>et al.</i>	39
A.3	Interpolazioni	39
B	Confronto tra il nostro modello sulla concentrazione di aminoacidi e quello di Kohanim <i>et al.</i>	40

Introduzione

L'utilizzo di colonie di batteri come *E.coli* come vettori per la produzione di proteine o gli studi riguardanti la resistenza dei batteri agli antibiotici necessitano di leggi chiare che descrivano tutti gli aspetti della crescita di una cellula. L'obiettivo di questo progetto è di cercare di descrivere uno degli aspetti che rientrano in questo problema, attraverso lo studio quantitativo dei fenomeni cellulari, sistemi ricchi di gradi di libertà, attraverso strumenti matematici forniti dalla fisica. Ciò è possibile anche grazie a nuovi strumenti di misura, in grado di seguire una popolazione di migliaia di cellule, tracciando i comportamenti di ognuna di esse in lunghi archi di tempo.

Il dati di questa tesi sono stati raccolti e messi a disposizione da Mia Panlilio del gruppo BSS (Biological and SoftSystems, Cavendish Laboratory) dell'Università di Cambridge. L'esperimento consiste nella crescita di una coltura di batteri *e. coli* in un apparato microfluidico che permette di seguire e tenere traccia della crescita di singole cellule [7]. Le cellule crescono in un ambiente di coltura minimale (M9+0.4% glucosio) per 7 ore, dopodiché vengono aggiunti dei caseoaminoacidi (M9 + 0.4% glucosio + CAA 0.5%) che aumentano il rate di traduzione e traslazione ribosomale con la conseguenza di una maggiore rapidità di crescita cellulare. Questo esperimento dà quindi la possibilità di avere dati sia di 2 stati stazionari a diverso ambiente di coltura, sia la possibilità di studiare come le cellule reagiscono e si adattano al cambio di nutrienti.

In particolare in questa tesi porremo l'attenzione sullo studio delle concentrazioni di ribosomi e proteine cataboliche nel proteoma della cellula, cercando di descrivere quantitativamente i loro valori a steady state e durante lo shift. Questo problema in letteratura è stato affrontato soprattutto su stati stazionari[3][8], delineando il formalismo settoriale (sezione 1) che rappresenta il framework alla base delle nostre analisi. In una prima parte (sezione 2) facciamo un'osservazione critica e approfondita dei dati che abbiamo a disposizione, cercando di definire nel modo migliore i proxies che ci restituiscano i valori delle grandezze che cerchiamo, dal momento che non possiamo misurare direttamente alcune grandezze microscopiche interne alla cellula.

Dopodiché mostreremo (sezione 4.2) come il modello FCR di Erickson *et al.*[3] sia basato su un'ipotesi che viene falsificata dai nostri dati e cercheremo

quindi nell'ultima parte di proporre dei modelli che rappresentino correttamente la crescita cellulare nei termini dei settori.

1 Modelli quantitativi di allocazione di risorse per la crescita cellulare

1.1 Le "growth laws" condensano la complessa dinamica di crescita cellulare in un modello a settori con pochi parametri

Il nostro studio parte dal framework introdotto da Hwa e Scott[8], esteso poi da Erickson e collaboratori al caso di transizioni tra stati di crescita diversi [3]. Questa teoria di tipo termodinamico semplifica la dinamica di crescita di un sistema cellulare, ricco di gradi di libertà, attraverso delle variabili macroscopiche legate alla crescita.

La prima variabile definita è quella che viene chiamata 'growth rate', indicata con λ . Solitamente con questa variabile viene definito il rate di crescita di 'bulk', ovvero dell'intera colonia di cellule, per cui abbiamo che in condizioni stazionarie di nutrienti (per esempio quando la concentrazione di nutrienti è mantenuta costante, come in un chemostato) la massa ed il volume totale crescono esponenzialmente con tale rate di crescita λ_M^* ¹

$$V(t) = V(0) \exp(\lambda_M^* t) .$$

Possiamo quindi definire

$$\lambda_M = \frac{1}{V} \frac{dV}{dt} .$$

Nel nostro caso, avendo a disposizione dati di singola cellula ed essendo a conoscenza del fatto che cellule di *E.coli* crescono in volume e massa in modo esponenziale [6, 10] tramite un rate di biosintesi λ_{bs} , possiamo affermare che il v di singola cellula segua esso stesso la legge

$$v(t) = v(0) \exp(\lambda_b^* t) .$$

Possiamo quindi definire

$$\lambda_{bs} = \frac{1}{v} \frac{dv}{dt} .$$

¹In tutta la tesi l'asterisco ad apice * sta ad indicare una grandezza relativa ad uno stato stazionario

Ogni cellula ha al suo interno una quantità enorme di proteine e molecole che interagiscono tra di loro in modo complesso. I determinanti della crescita però sono pochi, e sono legati a quanto la cellula riesce a bilanciare il flusso di nutrienti (con i quali costruisce i mattoni fondamentali che la formano) e il flusso di biosintesi (attraverso il quale la cellula "costruisce" se stessa a partire dai mattoni). Seguendo questo approccio macroscopico, per semplificarne lo studio, il proteoma (tutte le proteine prodotte da una cellula, che possiamo considerare la componente principale della biomassa) viene suddiviso in tre macrocategorie chiamate 'settori' in base al ruolo che svolgono. Nel settore ribosomale ϕ_R rientrano tutte quelle proteine che compongono i ribosomi, del settore catabolico ϕ_P fanno parte le proteine cataboliche non regolate della cellula (le proteine cataboliche sono responsabili della trasformazione di carbonio in amminoacidi) e dal settore di "housekeeping" ϕ_Q , è composto di proteine strutturali la cui frazione è regolata in modo da non cambiare in condizioni di crescita diversa. Si indica solitamente con ϕ_x la frazione di massa delle proteine del settore x rispetto alla massa totale del proteoma, $\phi_x = \frac{m_x}{m_{tot}}$. Ne segue che possiamo scrivere la relazione $\phi_R + \phi_P + \phi_Q = 1$. Questa relazione è di particolare importanza poiché in letteratura [8] è supposta l'esistenza del settore ϕ_Q che rimane costante al cambiare delle condizioni, quindi la somma

$$\phi_R + \phi_P = 1 - \phi_Q = \phi_{max} \quad (1.1)$$

con $\phi_{max} \approx 0.55$ [8], ovvero il settore R e P sono complementari.

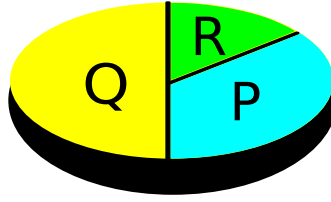


Figure 1.1: Illustrazione grafica dei settori del proteoma. Ogni settore rappresenta tipi diversi di proteine. I membri del settore R sostengono la biosintesi, mentre gli altri settori sono prodotti passivamente. Q rimane costante durante lo shift mentre P ed R cambiano in modo complementare

Uno dei più importanti risultati in questo campo è sicuramente la cosiddetta "prima legge" (first growth law), formulata da Hwa *et al.*[8], che lega il settore ϕ_R al rate di crescita λ in stati stazionari, *a prescindere da ogni dettaglio molecolare*. Difatti si riscontra una proporzionalità lineare tra queste due grandezze, esprimibile nella forma

$$\phi_R^* = \phi_0 + \frac{\lambda^*}{\gamma}, \quad (1.2)$$

dove ϕ_0 e γ sono due parametri che indicano rispettivamente intercetta e inverso del coefficiente angolare della retta di proporzionalità, individuati dai dati.

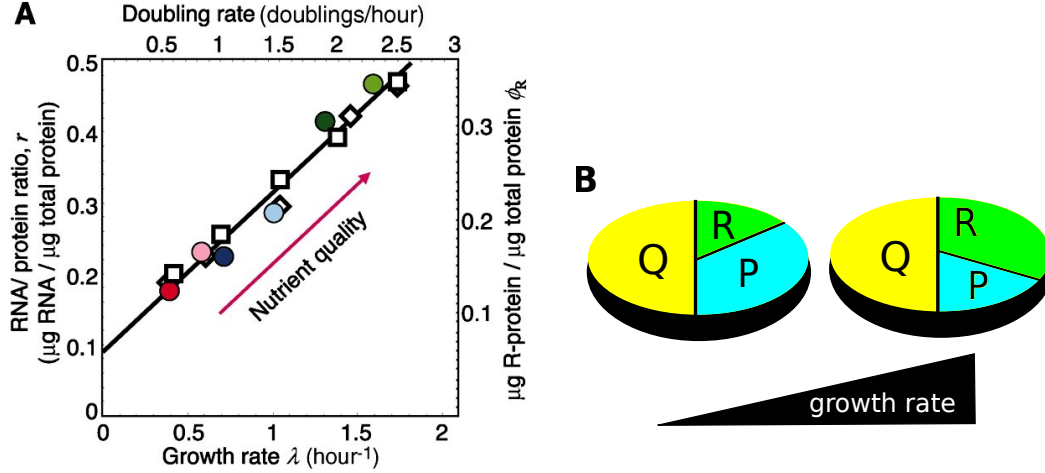


Figure 1.2: **A.** I dati a steady state di Hwa, Scott *et al.*[8] seguono la prima growth law lineare, Eq. (1.2). **B.** Mentre il settore Q rimane costante, un aumento di λ corrisponde ad un aumento di ϕ_R (cioè dei ribosomi) a discapito di ϕ_P .

È noto in letteratura che la variazione del numero di ribosomi al variare dei nutrienti è regolata nei batteri principalmente dalla molecola ppGpp (Guanosin tetrafosfato). ppGpp è in grado di operare sensing del livello di amminoacidi disponibili nella cellula, favorendo la produzione di proteine del settore R quando la disponibilità di amminoacidi è alta, inibendola quando è bassa [1]. Si può comprendere qualitativamente questa regolazione pensando che alcune delle proteine P che vengono prodotte sono necessarie al processo catabolico che trasforma i nutrienti a base di carbonio in amminoacidi, quindi quando ci troviamo in una situazione di nutrienti di scarsa qualità abbiamo bisogno di un numero maggiore di proteine cataboliche che allargano quindi il settore P a discapito di R (vale il viceversa quando aumentiamo la qualità dei nutrienti) come possiamo vedere in fig.1.2(A).

La regolazione attraverso il ppGpp realizza l'ottimo di questo trade-off. Il modo in cui agisce il ppGpp passa per la *trascrizione* dei geni ribosomali, cioè dalla produzione del RNA messaggero per questi geni. Infatti il ppGpp si lega alla RNA polimerasi (che legge i geni sul DNA e produce l'RNA corrispondente) e inibisce la lettura dei geni ribosomali (che contengono una sequenza specifica). I livelli di ppGpp aumentano quando aumentano i tRNA "scarichi" (che non trasportano codoni), cioè in presenza di pochi nutrienti, grazie a una proteina specifica (RelA) che ne catalizza la produzione quando è legata a un tRNA scarico.

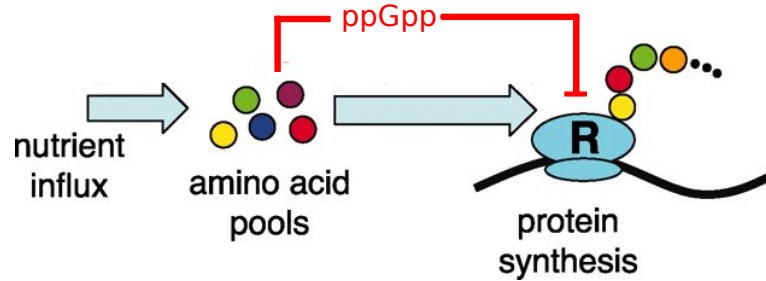


Figure 1.3: Illustrazione semplificata del processo di sintesi proteica [8]. La molecola ppGpp regola la produzione di ribosomi a partire dai livelli di aminoacidi a disposizione.

1.2 Il modello di regolazione controllata da flussi (FCR) di Erickson si propone di spiegare le transazioni tra stati di crescita cellulare

Lo studio delle variazioni dei settori all'interno di una cellula al variare dei nutrienti parte dalla definizione di J_C , ovvero il *flusso di carbonio* che viene estratto dai substrati di nutrimenti, che nell'esperimento di Erickson sono degli zuccheri, ricchi di tale sostanza. Questo carbonio viene trasformato in amminoacidi, gli ingredienti principali che vengono utilizzati dai ribosomi per sintetizzare le proteine. Da qui una delle ipotesi principali del modello FCR: il flusso di nutrimenti provenienti dalle sorgenti di carbonio J_C deve essere bilanciato con quello di sintesi proteica

$$J_R = \alpha J_C \quad (1.3)$$

con α costante di conversione tra carbonio e proteine. È importante sottolineare che viene precisato, però, che questa ipotesi è valida in scale temporali 'coarse grained' (traducibile con 'a grana grossa') che hanno degli step temporali di ~ 15 min. Questa ipotesi, riassumendo, ci dice che se non raffiniamo troppo la definizione temporale delle misure, possiamo pensare i due flussi come sempre bilanciati, considerando i meccanismi cellulari che intercorrono tra l'assorbimento del carbonio (J_C) e la sintesi proteica (J_R) sulla stessa scala temporale a grana grossa.

Il modello FCR di Erickson, cerca di descrivere partendo proprio dal bilancio di flusso di nutrimenti in entrata e di sintesi proteica, le transazioni tra stati di crescita cellulare. Per descrivere la risposta dei ribosomi ai cambiamenti di

nutrimenti definisce:

$$\sigma = \frac{J_R}{M_{rb}} = \frac{\lambda}{\phi_R} \quad (1.4)$$

ovvero il rate di sintesi proteica di un ribosoma (è definito infatti come il flusso di proteine sintetizzate sulla massa totale dei ribosomi). Ricordando poi che le proteine sintetizzate (tralasciando quelle di *housekeeping*) possono essere proteine ribosomiche (settore R) o costitutive (settore P) è importante capire qual è la ripartizione delle proteine sintetizzate tra questi due settori. Abbiamo già spiegato qualitativamente che la proteina ppGpp regola in base al numero di aminoacidi disponibili se le nuove proteine sintetizzate saranno R o P. Introduciamo per descrivere quantitativamente questo processo le due funzioni di ripartizione χ_R e χ_P tali che:

$$\begin{cases} \frac{dM_{rb}}{dt} = \chi_R(t)J_R \\ \frac{dM_{cat}}{dt} = \chi_P(t)J_R \end{cases} \quad (1.5)$$

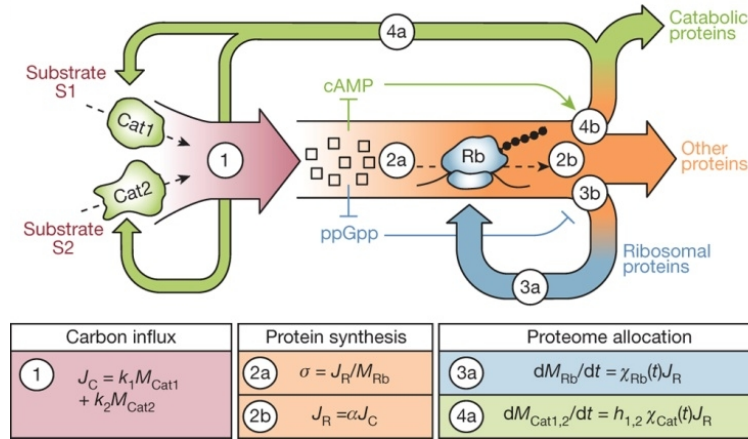


Figure 1.4: Illustrazione grafica del modello FCR. La freccia 1 indica la trasformazione del carbonio estratto dai substrati in amminoacidi, che vengono assorbiti dai ribosomi per la sintesi di proteine ribosomali, cataboliche e di housekeeping, secondo le equazioni descritte.

L'equazione principale del modello FCR di Erickson è una equazione differenziale per σ

$$\frac{d\sigma}{dt} = \mu_f \left[\frac{1 - \sigma(t)/\sigma_f}{1 - \sigma(t)/\gamma} \right] \quad (1.6)$$

che nasce però dall'ipotesi fondamentale che verrà discussa successivamente in questa tesi: il plot di χ vs σ nei dati di Erickson in diversi steady state (ogni punto rappresenta un χ_i^* vs σ_i^*) sembra mostrare una curva.

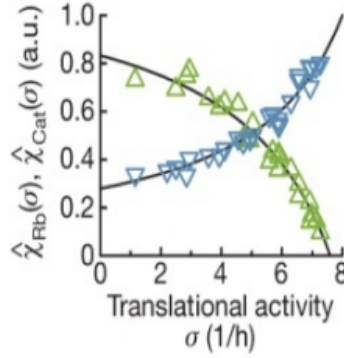


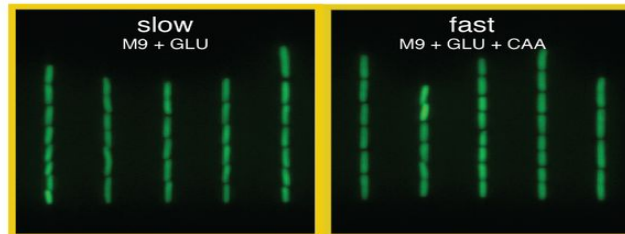
Figure 1.5: Gli steady state di χ e σ individuano una curva nei dati di Erickson[3]. I punti blu rappresentano χ_R , quelli verdi χ_P .

Perciò viene ipotizzato nel modello che $\chi_R(t) = \hat{\chi}_R(\sigma(t))$ quindi che χ_R possa essere espresso come funzione di σ per ogni tempo, anche fuori dagli steady state e che il sistema evolva lungo la curva individuata in fig(1.5) seguendo degli stati metastabili.

2 Esperimenti in chip microfluidici permettono di estrarre la dinamica dei settori

Abbiamo introdotto fino ad ora concetti e modelli in letteratura che mirano a descrivere quantitativamente la dinamica di crescita cellulare. È d'obbligo a questo punto, lasciando un attimo in sospeso l'analisi dei modelli, introdurre anche i dati coinvolti. Questo lavoro parte da esperimenti di shift di nutrienti svolti da M. Panlilio all'Università di Cambridge (Gruppo Cicuta) [7].

Figure 2.1: Frame della crescita cellulare nell'esperimento di Mia Panlilio



L'esperimento consiste in un chip microfluidico dove i batteri possono crescere in canaletti micrometrici in condizioni molto controllate, e sono soggetti a un flusso continuo di nutrienti, che rimuove anche tutte le cellule che escono dai

canaletti. In questo "microchemostato", osservato al microscopio, la biomassa rimane costante, ed è possibile seguire dinamicamente le proprietà di cellule singole, anche se in questo lavoro ci occuperemo solo delle medie macroscopiche.

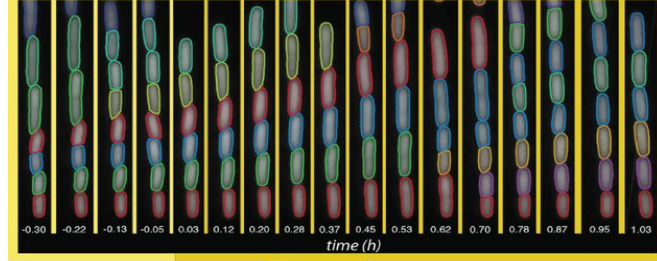
I frame del sistema vengono presi ogni 5 minuti per un arco totale di 700 minuti (da $t = -300min$ a $t = 400min$ dove $t = 0$ rappresenta il momento dello shift). I canaletti sono organizzati in 35 FOV (field of view), ovvero le diverse posizioni che assume la macchina che raccoglie i frame poichè non riesce a 'fotografare' l'intero sistema di canaletti contemporaneamente. Ogni misurazione viene ripetuta per più repliche, in diversi giorni.

In questo tipo di setup la concentrazione di ribosomi ϕ_R non può essere misurata direttamente in una colonia di cellule, perciò se siamo interessati ai settori necessitiamo di alcuni indicatori che ci diano indirettamente le misure di cui abbiamo bisogno. In letteratura sono stati utilizzati per la maggior parte dei casi dei dati di bulk, ovvero dei dati relativi alla popolazione intera di cellule, utilizzando come proxy ribosomale la quantità di RNA sulla quantità totale di proteine nelle cellule [3, 8].

Nel nostro caso, avendo un apparato sperimentale che permette misure di singole cellule, vi era bisogno di indicatori che ci restituissero le misure per ogni cellula. Perciò è stata inserita sui promoter la sequenza relativa alla proteina GFP (una proteina che conferisce fluorescenza alla cellula) nel DNA batterico. In pratica un promoter (che sia ribosomale o costitutivo) è una regione del DNA batterico vicina ad una particolare sequenza che contiene il codice per trascrivere un gene. L'RNA polimerasi, responsabile di tale trascrizione, si lega al promoter e da qui legge il codice generando la proteina corrispondente.

Perciò la proteina GFP dovrebbe essere prodotta allo stesso rate di produzione dei ribosomi, quando il suo codice è "innestato" sul promotore ribosomale P1 (soggetto agli effetti di ppGpp) o allo stesso rate delle proteine costitutive se innestato sul promotore costitutivo P5 (non soggetto alla regolazione attraverso ppGpp). A meno di effetti del GFP sul proteoma, ci aspettiamo quindi che la fluorescenza nelle cellule sia proporzionale al numero (e quindi alla massa) dei ribosomi, nel caso stessimo osservando i dati relativi a un promoter P1, delle proteine costitutive nel caso stessimo osservando i dati relativi a un promotore P5. Infine è rilevante stabilire dove è stato inserito il promotore riportato da GFP, perchè la produzione di un gene dipende dal numero di copie del DNA, e siccome il DNA viene replicato in circa 40 minuti sempre nello stesso ordine (da una "origine" a un "terminus") geni vicini all'origine hanno tipicamente più copie del DNA di geni vicini al terminus. I ribosomi, nei batteri, sono sempre vicini all'origine, mentre i promotori costitutivi non hanno un bias per posizioni genomiche specifiche.

Figure 2.2: La proteina GFP conferisce alle cellule di *E.coli* una certa fluorescenza che viene misurata da un software attraverso i frame dei canaletti



Negli esperimenti effettuati dai nostri collaboratori, promotori ribosomali (P1) e costitutivi (P5) sono stati inseriti in diverse posizioni, e in particolare vicino all'origine e vicino al terminus (P1longori , P1ori, P1ter, P1longter, P5ori, P5ter)². La prima analisi che dovremo fare sul nostro dataset mira a comprendere, a partire dalle misure di fluorescenza e volume, quali siano tra quelli a disposizione i migliori proxy per ϕ_R , ϕ_P e λ .

²Il termine *long* rappresenta la presenza dell'intera sequenza di controllo

Risultati

3 I dati a disposizione permettono di estrarre tutti i parametri rilevanti per le leggi di crescita e le transizioni

Questo capitolo descrive i risultati principali ottenuti analizzando i dati dei collaboratori di Cambridge alla luce di modelli dinamici a settori per la biosintesi. Il lavoro svolto è può essere definito come "model building". In una prima fase ho lavorato sui dati per ottenere delle osservabili affidabili e collegate alle osservabili principali considerate dai modelli a settori. In una seconda fare ho confrontato i dati con vari modelli allo scopo di validarli/falsificarli, procedendo in base alle analisi a proporre modelli e ingredienti che potessero descrivere in modo più accurato i dati.

3.1 Il promotore costitutivo in ter e ribosomale in ori rappresentano bene i rispettivi settori del proteoma.

I grafici della fluorescenza F prodotta dai reporter GFP possono darci una prima indicazione della bontà del promoter utilizzato come proxy per i settori. Difatti utilizzando la definizione (1.1) e assumendo (come ragionevole aspettarsi in condizioni lontane dalla saturazione della fluorescenza) che la fluorescenza totale sia proporzionale al numero di proteine GFP espresse, a sua volta proporzionale al numero (e quindi alla massa) di ribosomi o proteine cataboliche ($N_{P1} \sim N_{\text{ribosomi}}$ e $N_{P5} \sim N_{\text{proteine-cataboliche}}$), possiamo scrivere che,

$$\begin{aligned} F_{P1} &= C_1 M_R & F_{P5} &= C_2 M_P \\ , \end{aligned}$$

dividendo per il volume e considerando costante la densità si ottiene

$$\frac{F_{P1}}{V} = \rho C_1 \frac{M_R}{M} \qquad \frac{F_{P5}}{V} = \rho C_2 \frac{M_P}{M}$$

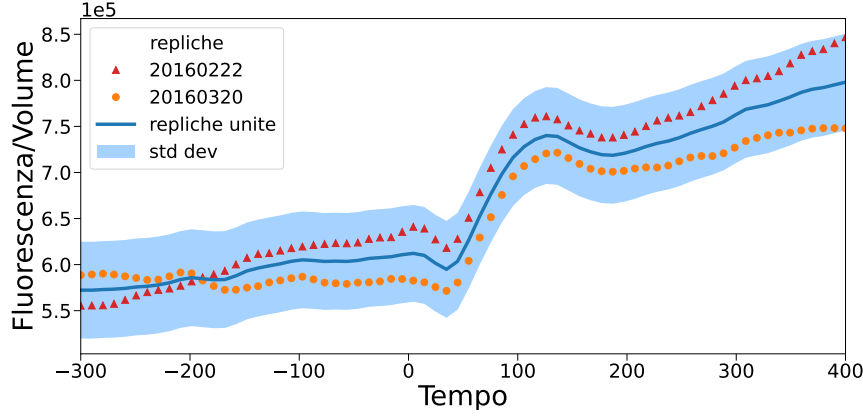


Figure 3.1: Le due repliche biologiche dell'esperimento con il promotore P1ori sono coerenti tra di loro. Le repliche rientrano in una deviazione standard l'una dall'altra. Inoltre lo shift passa correttamente da uno stato a minor concentrazione di ribosomi a uno a maggiore concentrazione, ma con il problema che il secondo stato stazionario non è ben definito dato che la concentrazione continua a crescere. Probabilmente questo problema (riscontrato in un esperimento più che nell'altro) è dovuto a flusso non perfetto dei nutrienti nel microchemostato a tempi lunghi.

da cui infine

$$\frac{F_{P1}}{V} = A_1 \phi_R \qquad \frac{F_{P5}}{V} = A_2 \phi_P . \quad (3.1)$$

In altre parole le concentrazioni dei due promotori sono indicatori delle frazioni dei relativi settori.

Ne segue che la media mobile del rapporto fluorescenza/volume in singole cellule del nostro esperimento è proporzionale tramite una costante moltiplicativa da determinare all'andamento delle frazioni settori. Questo implica che un buon proxy deve individuare due steady state ben definiti nel grafico F/V , uno pre e uno post shift di nutrienti. Nel caso di P1 (settore ribosomale) dobbiamo trovare che la frazione occupata dal settore nel secondo stato stazionario è maggiore che nel primo, e naturalmente il contrario nel caso di P5. Tutti i grafici della media del rapporto fluorescenza/volume sono ottenuti tramite una media mobile sui tempi, misurati rispetto all'evento di shift dei nutrienti, come spiegato in dettaglio in appendice A.

Per quanto riguarda il settore ϕ_R , P1longori rispetta l'aspettativa e si conferma il miglior proxy ribosomale tra quelli a disposizione (Fig. 3.1 e 3.2). Per il settore ϕ_P (Fig. 3.4 e 3.3) ci aspettiamo una diminuzione della concentrazione di fluorescenza, sempre individuando nettamente i due stati stazionari. I dati non

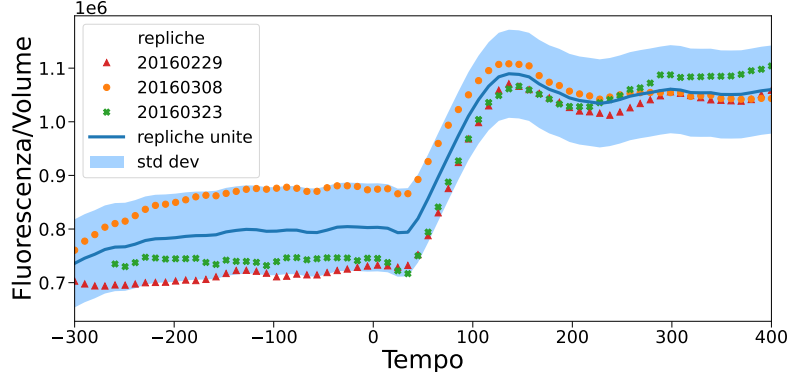


Figure 3.2: Le repliche biologiche P1longori appaiono come un buon proxy per il settore ribosomale. Due repliche per questo promotore sono quasi identiche. La replica 0308 per quanto meno coerente con le altre due, rientra in una deviazione standard. Inoltre le tre repliche mostrano correttamente i due stati stazionari e lo shift che li collega. Infine il promotore P1long contiene tutte le sequenze di regolazione di un vero promotore ribosomale e si può quindi considerare un reporter migliore dal punto di vista biologico.

forniscono per ϕ_P un proxy altrettanto buono come lo è P1longori per ϕ_R . P5ori è difatti inutilizzabile poiché l'andamento previsto per il settore ha una forma molto diversa da quella attesa, probabilmente dovuta al fatto che i promotori del settore p non hanno un bias verso l'origine (Fig. 3.3). P5ter invece ha una forma più vicina a quella attesa, sebbene gli esperimenti non individuino nitidamente il secondo steady state, probabilmente a causa di problemi nei flussi di nutrienti a tempi lunghi dell'esperimento (Fig. 3.4).

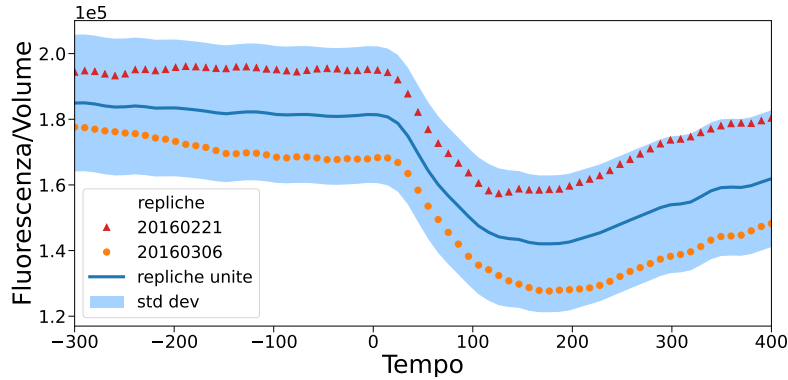


Figure 3.3: Per P5ter abbiamo due repliche abbastanza coerenti, rientrano in una dev std. Dopodiché il grafico di P5ter sembra avere una forma più accettabile, in quanto è ben visibile il downshift della fluorescenza, ma anche in questo caso non viene raggiunto in modo ben definito il secondo steady state.

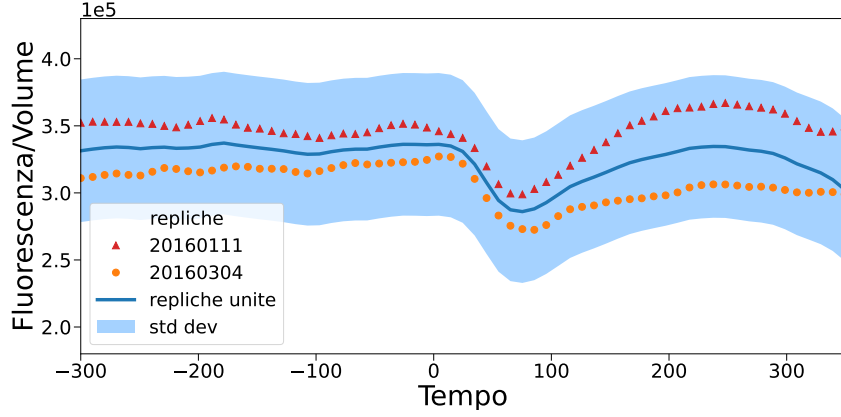


Figure 3.4: Per quanto P5ori presenti delle repliche coerenti, ha diversi problemi: come prima cosa una replica che avevamo a disposizione è stata eliminata poichè era poco coerente con le altre, non rientrando in una dev std che è comunque più alta in percentuale rispetto alle altre strain. Inoltre dal grafico di P5ori capiamo che non può essere utilizzato come proxy per ϕ_P . Difatti per quanto si veda un abbassamento del livello di fluorescenza dopo lo shift a $t=0$, non abbiamo un assestamento su un valore costante ma anzi abbiamo che la concentrazione di fluorescenza cresce.

Vogliamo ora utilizzare i dati per eliminare l'incertezza sulle costanti di proporzionalità tra fluorescenza e frazioni dei settori del proteoma. Il primo tentativo di definire le frazioni dei settori dai proxy sfrutta le espressioni di ϕ_R e ϕ_P trovate dalla (3.1), inserendole in (1.1):

$$\frac{1}{A1} \frac{F_{P1}}{V} + \frac{1}{A2} \frac{F_{P5}}{V} = \phi_{max} \quad (3.2)$$

Abbiamo visto come P1longori e P5ter si presentino come i migliori proxy rispettivamente di ϕ_R e ϕ_P , quindi sostituiamo nella Eq. (3.2) le serie temporali $\frac{F_{P1longori}}{V}$ e $\frac{F_{P5ter}}{V}$. Attraverso un fit³ si ottengono le costanti A1 e A2 (Fig. 3.5). Il fatto che si riescano ad ottenere due costanti A1 e A2 tali per cui la definizione (3.2) è rispettata abbastanza bene lungo tutto l'arco temporale delle misure è garanzia del fatto che -pur non essendo P5ter un proxy ottimo per ϕ_P - le strain scelte possono essere considerate buoni proxy dei settori.

³A questo fine ho utilizzato la funzione `scipy.optimize.curve_fit`

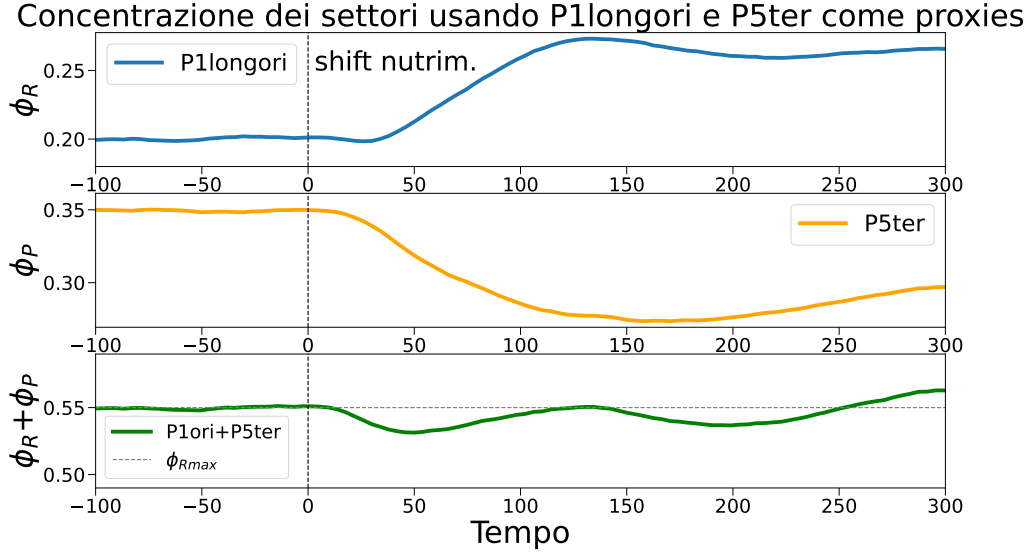


Figure 3.5: I settori del proteome definiti dalla fluorescenza attraverso la definizione in Eq. 3.2 sommano coerentemente a una costante lungo l'esperimento. La somma di ϕ_P e ϕ_R rimane abbastanza costante attorno al valore ϕ_{max} quasi per tutta la durata temporale delle misure, alzandosi leggermente da tale valore costante per $t > 200$ a causa della non stabilità del valore di ϕ_P già discussa in precedenza

3.2 Il reporter ribosomale P1longori è utilizzabile da solo come proxy 'sigle strain' per ϕ_R e ϕ_P grazie alla growth law

Sebbene la definizione precedente di settori porti ad un risultato abbastanza soddisfacente, vi sono ancora dei problemi da affrontare. In primis, la definizione (3.2) funziona attraverso l'utilizzo di due esperimenti diversi contemporaneamente, quindi il valore di ϕ_R , ad esempio, non dipende solamente dalla strain P1 che scegliamo per rappresentarlo, ma anche dalla strain P5 complementare che inseriamo nella definizione. Inoltre abbiamo già discusso che l'andamento temporale medio del rapporto fluorescenza / volume (e quindi di ϕ_P) di P5ter nelle repliche sperimentali a nostra disposizione non individua in modo nitido il secondo steady state, quindi vorremo cercare un modo per evitare il suo utilizzo.

Per risolvere questi problemi ci viene in aiuto la growth law (1.2), che ci permette di definire ϕ_R dalla singola strain di P1longori, in quanto, sempre considerando valida la proporzionalità discussa in (3.1) tra Fluo/Volume e ϕ_R , pos-

siamo scrivere:

$$B \frac{Fluo_{p1}}{V} = \phi_R^* = \phi_0 + \frac{\lambda^*}{\gamma} \quad (3.3)$$

dove γ e ϕ_0 sono due costanti misurate in [3], rispettivamente del valore di $\gamma \simeq 0.183[\text{min}^{-1}]$ e $\phi_0 \simeq 0.049$. Nell'equazione (3.3) utilizziamo il growth rate λ di P5ter per i motivi che verranno specificati nella prossima sezione. Tramite la growth law (3.3) ricaviamo 2 parametri $B_{1,2}$ rispettivamente dai dati del primo e del secondo steady state di λ e Fluo/Volume di P1longori.

$$B1 = \left(\frac{V}{Fluo_{p1}} \right)_1 \left(\phi_0 + \frac{\lambda_1^*}{\gamma} \right) \quad B2 = \left(\frac{V}{Fluo_{p1}} \right)_2 \left(\phi_0 + \frac{\lambda_2^*}{\gamma} \right)$$

A questo punto posso calcolare:

$$\phi_R = B \frac{Fluo_{P1}}{V} \quad (3.4)$$

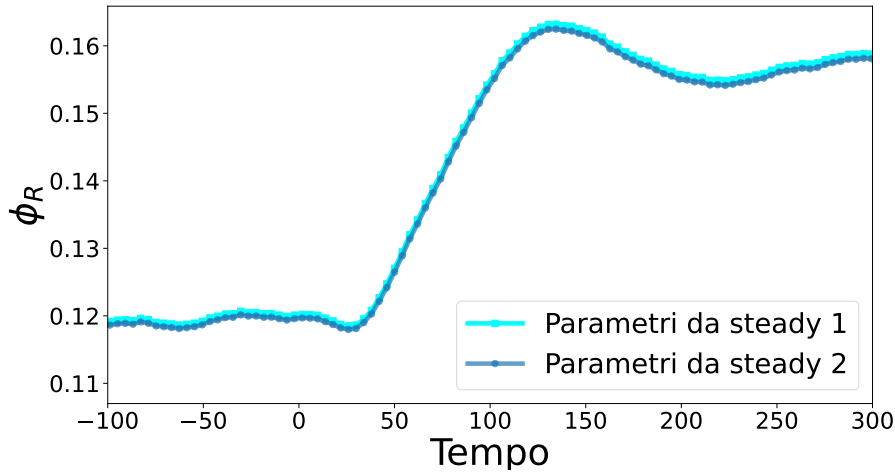


Figure 3.6: Le costanti $B_{1,2}$ ricavate dai dati dei due steady state sono pressoché identiche, sintomo di un ottimo accordo tra i nostri dati e la growth law. Inoltre sono ben individuati lo steady state pre e post shift.

Figura 3.6 mostra che le costanti $B_{1,2}$ ricavate dai dati dei due steady state sono pressoché identiche. Di conseguenza P1longori si rivela il miglior proxy per il settore ribosomale non solo poiché è l'unico che individua nitidamente i due steady state, ma anche poiché è quello che meglio rispetta la growth law, difatti le costanti che legano fluorescenza e ϕ_R sono identiche se calcolate sul primo o sul

secondo steady state (da ora in poi verrà semplicemente indicato il parametro B, media di B1 e B2).

Proprio per questi motivi e per i problemi già constatati riguardo P5ter come proxy costitutivo decidiamo di utilizzare come valore di per il settore P $\phi_P = \phi_{max} - \phi_R$ con ϕ_R calcolato da (3.4).

3.3 Lo strain con il miglior proxy per il rate di crescita è P5ter

Abbiamo quindi definito il settore ribosomale ϕ_R tramite la growth law utilizzando la fluorescenza di P1longori, ma che λ abbiamo utilizzato? La growth law suggerirebbe l'utilizzo del growth rate della stessa strain da cui si ricava la fluorescenza, per una maggior coerenza. Nel nostro caso, però, abbiamo visto che i dati dei rate di crescita di diverse strain, in particolare quello di P1longori, presentavano dei problemi e delle incongruenze (mostrate in seguito). La visione standard [8] è che GFP fa parte del settore P, e quindi in uno stato stazionario dovrebbe ridurre la frazione del proteoma disponibile per il settore ribosomale. Questo non è verificato nei nostri dati (Figura 3.7) dove le strain P1ori, che esprimono più GFP essendo il promoter vicino all'origine di replicazione mostrano un growth rate più alto, e la strain P5ter, che esprime meno GFP di tutte, mostra i growth rate più bassi. Durante un upshift, è stato riportato che la cellula ha un costo di adattamento per compensare la presenza del GFP, che aumenta con la sua concentrazione [9]. Anche questo non sembra essere del tutto il caso per i nostri dati, essendo anche nel transiente il growth rate più basso per la strain p5ter e più alto nel caso delle strain P1ori. In base a entrambe le teorie il proxy migliore dei settori dovrebbe essere P5ter, perchè la bassa espressione minimizza l'impatto del GFP sul proteoma, ma dati i problemi nei dati la scelta risulta difficile. Per risolvere al meglio questo problema dati i limiti dei dati abbiamo analizzato in dettaglio l'andamento del growth rate per ogni strain per trovare quella che meglio è coerente con il framework che stiamo utilizzando, considerando sia le variazioni all'interno del chip microfluidico, sia quelle temporali, sia quelle tra repliche dell'esperimento.

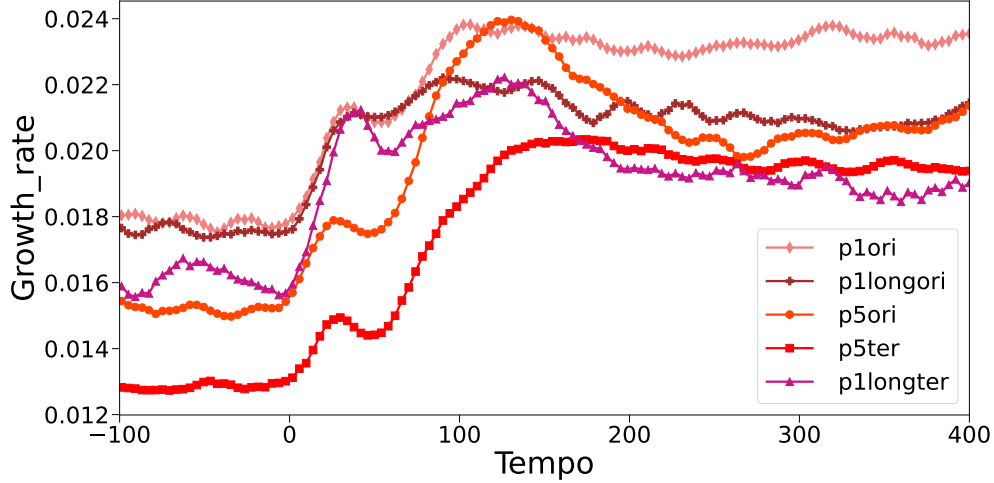


Figure 3.7: Le serie temporali di growth rate deviano dalle aspettative. Le serie sono ottenute come spiegato in appendice A. Sebbene ci siano delle somiglianze importanti tra i grafici delle varie strain, come il primo piccolo overshoot subito dopo lo shift di nutrimenti, sono evidenti sia l'incoerenza che si riscontra tra una strain e l'altra, sia le specificità di alcune strain. **(i)** Sia P5ori, sia P1lognter non rispondono al comportamento atteso poiché presentano un picco attorno a $t \simeq 120min$ per poi assestarsi su un valore inferiore mentre ci aspettiamo un grafico che passi, attraverso uno shift, da un primo stato stazionario ad un altro maggiore, ben individuati e nitidi, senza che ci sia un picco ed un abbassamento. **(ii)** In P1longori riscontriamo invece un problema diverso: la differenza del valore tra il secondo steady state e il primo è troppo bassa, il growth rate cresce troppo poco, e questo ci porta ad una conclusione biofisicamente incoerente quando andiamo a calcolare σ da (1.4) poiché il valore di σ nel secondo steady state risulta essere minore del primo ma noi sappiamo che se aumentiamo i nutrimenti il rate di produzione dei ribosomi deve necessariamente crescere. **(iii)** Le due strain che presentano i grafici migliori sono indubbiamente P5ter e P1ori, quindi dovremmo andare a capire quale utilizzare.

Per avere una maggiore chiarezza nello studio di questi dati abbiamo deciso di analizzare le varie repliche delle strain che abbiamo preso in considerazione (P5ter, P1ori e P1longori) suddividendo ogni replica in sottopopolazioni composte da gruppi di 5 FOV (field of view, vedi Sec.2) ognuna, per vedere se sono individuabili dei trend sistematici di errore che ci diano una risposta sulla poca coerenza riscontrata nei grafici di λ .

3.3.1 Dati di growth rate di P1longori suddivisi per FOV evidenziano eterogeneità nella crescita

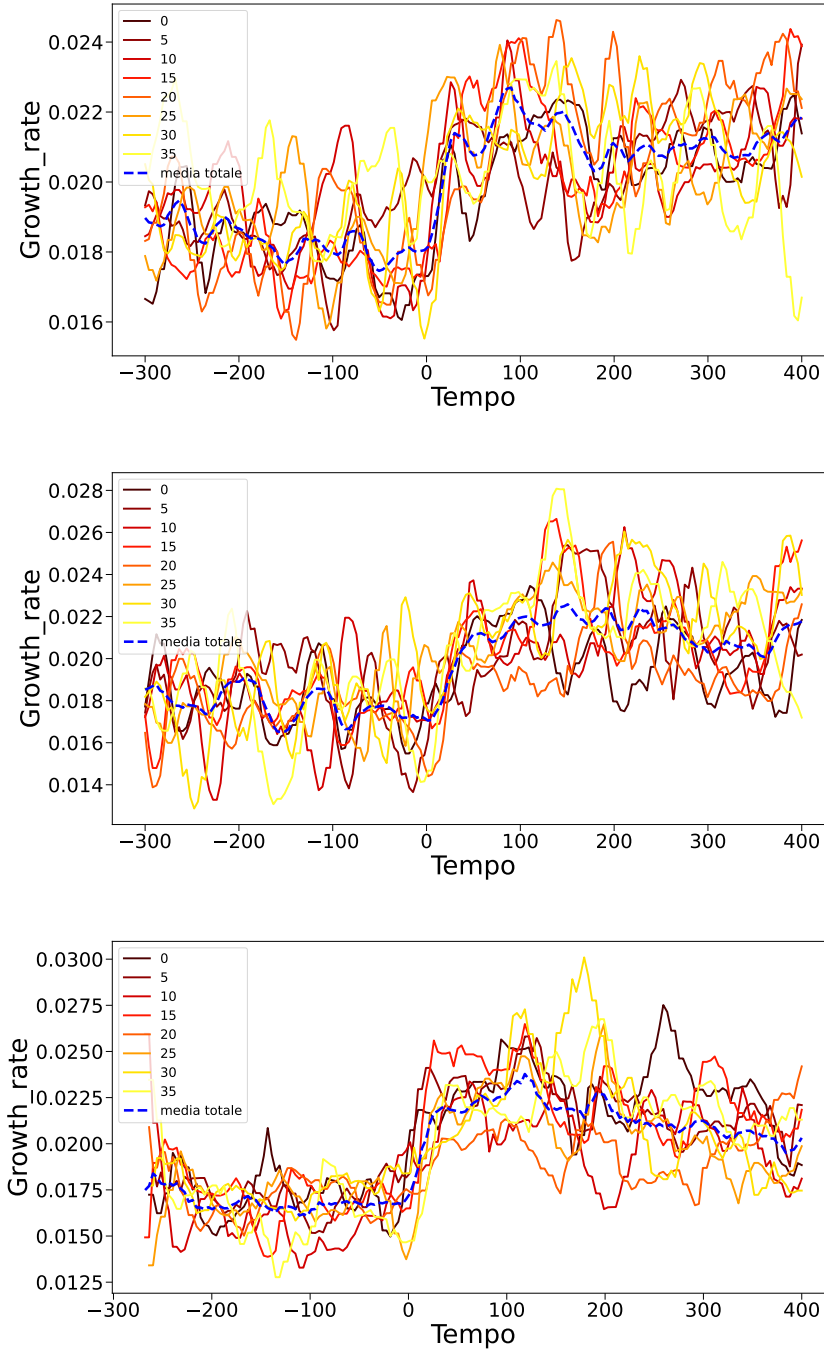


Figure 3.8: I dati del growth rate di P1longori mostrano forti oscillazioni nei singoli FOV, per ogni replica (ogni replica è mostrata in un pannello diverso). Non riusciamo a riconoscere un trend di errore sistematico in dipendenza dal FOV o da altri fattori ma osserviamo la presenza di oscillazioni ampie, casuali e non controllate, dovute forse a qualche errore nella presa delle misure (ad esempio la macchina può aver perso il fuoco nelle immagini, etc.) che rendono questi dati poco coerenti tra di loro e inutilizzabili.

Sebbene il dato del growth rate venga estrapolato dal calcolo di una derivata temporale, che può causare molto rumore, in questo caso siamo davanti a dei dati che non possono essere in alcun modo considerati buoni poiché presentano molta incoerenza tra diversi FOV e delle oscillazioni attorno al valore medio che arrivano anche a $\pm 20\%$. Notiamo che parte delle oscillazioni osservate potrebbe anche avere origine biologica nelle eterogeneità del comportamento tra genealogie diverse. Questa ipotesi è interessante ma va oltre gli scopi di questo lavoro. Inoltre non la riteniamo completamente attendibile, rispetto a problemi di fuoco del microscopio e attendibilità delle derivate, dato che le oscillazioni non sono riscontrate nell'altra osservabile misurata sulla stessa strain, la concentrazione di GFP (Fig. 3.9).

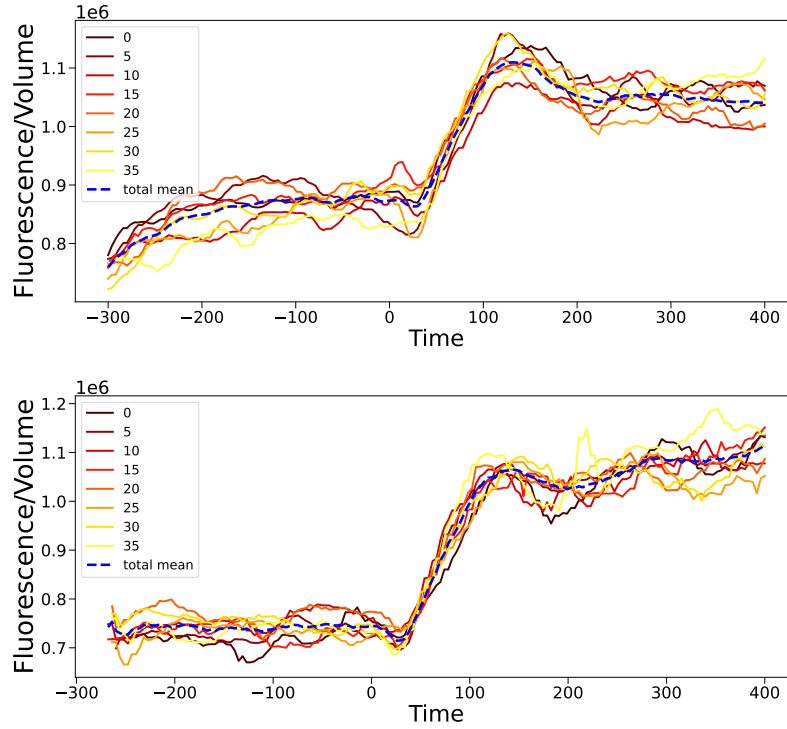


Figure 3.9: I dati della fluorescenza/volume di P1longori mostrano un'ottima coerenza tra i vari FOV, tra varie strain (una per ogni pannello) e con il framework, a differenza dei dati di growth rate. Difatti abbiamo già mostrato come P1longori risulti essere un ottimo proxy per ϕ_R . Questo ci suggerisce che i problemi riscontrati nei dati di growth rate possano essere imputabili più a errori di misura o affidabilità delle derivate piuttosto che ad un'origine biologica

3.3.2 Dati di growth rate di P1ori suddivisi per FOV evidenziano eterogeneità nella crescita.

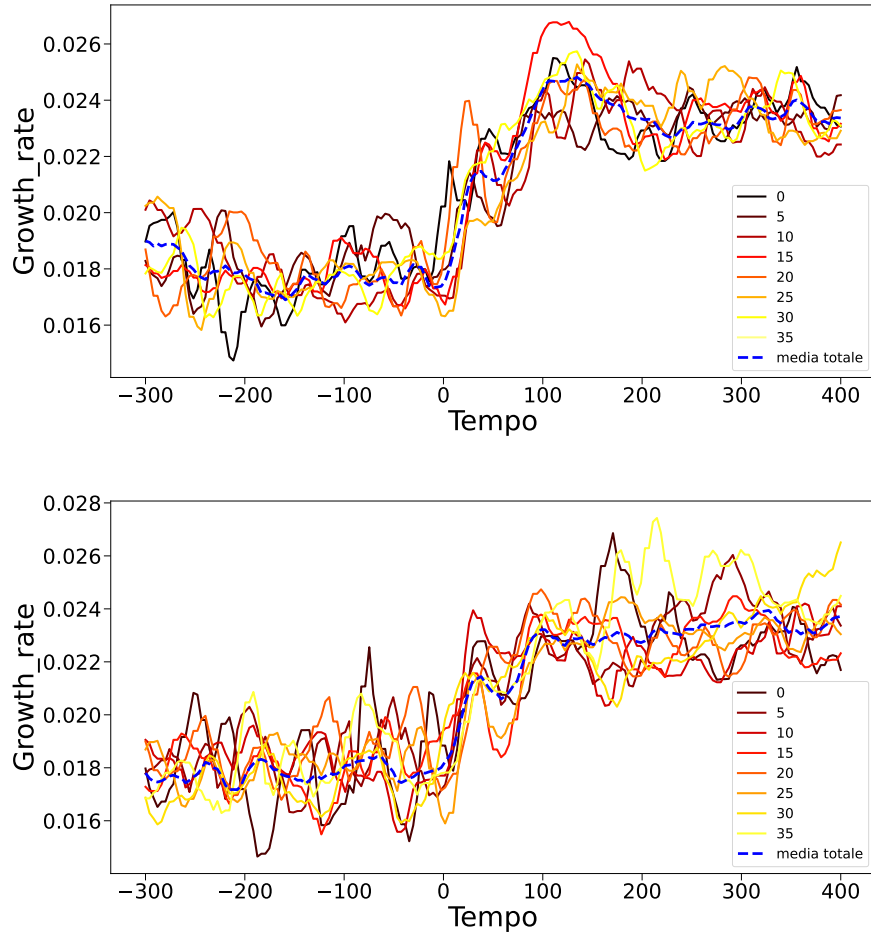


Figure 3.10: I dati del growth rate di P1ori mostrano caratteristiche simili a quelli di P1longori: forti oscillazioni nei FOV, per ogni replica. Le oscillazioni sono ampie e appaiono casuali, sebbene di minore ampiezza di quelle di P1longori. L'origine del problema potrebbe essere qualche errore nella presa dati.

I grafici di growth rate di P1ori (Fig. 3.10) non sono sicuramente tanto oscillanti e incoerenti quanto quelli di P1longori ma comunque non hanno una buona regolarità attorno al valore medio, soprattutto in alcune sottopopolazioni abbiamo delle oscillazioni importanti.

3.3.3 Dati di growth rate di P5ter suddivisi per FOV confermano tale strain come miglior proxy

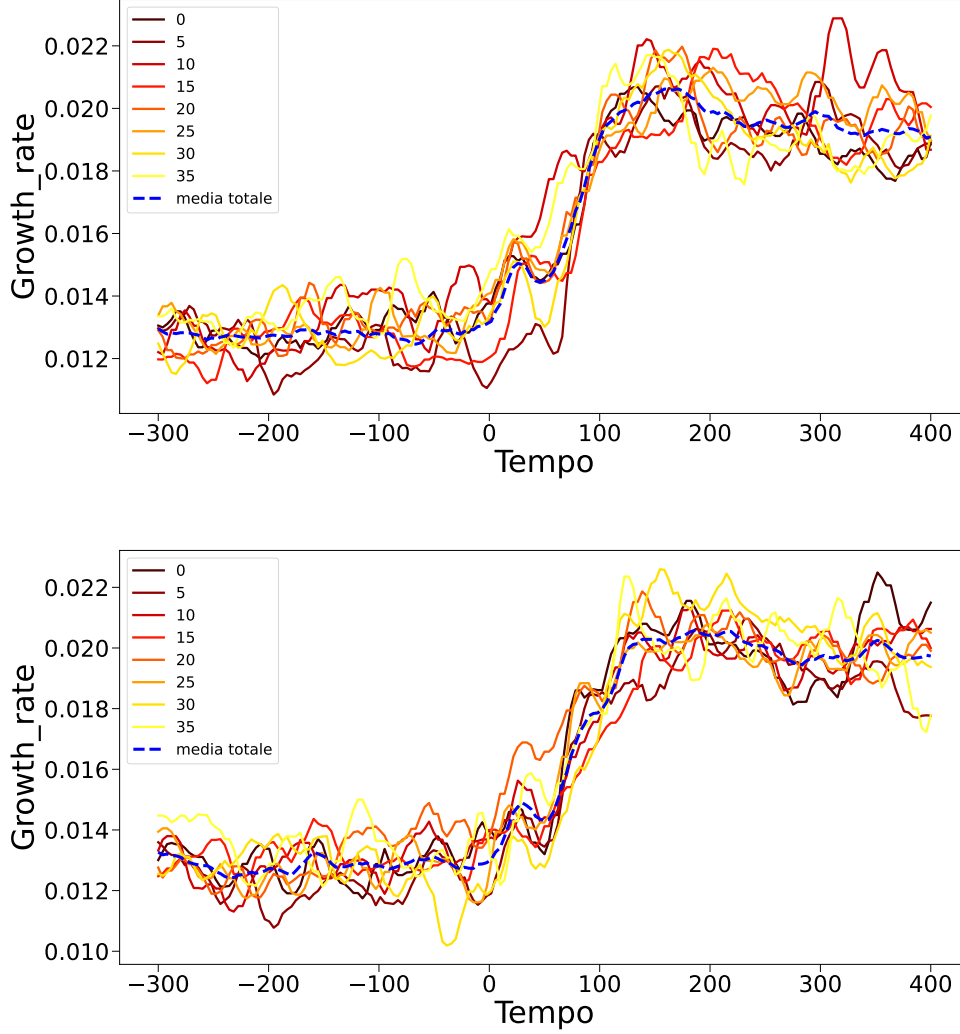


Figure 3.11: A differenza delle due strain viste in precedenza, per P5ter abbiamo dei grafici più coerenti tra i diversi FOV e meno oscillanti, soprattutto nella replica 0306.

Per la sua coerenza temporale, tra FOV e repliche biologiche, P5ter si dimostra il miglior proxy per quanto riguarda il growth rate (Fig. 3.11). Tra le tre strain che abbiamo preso in considerazione essa è quella che presenta una maggior coerenza tra FOV, con i dati meno oscillanti. Inoltre, come abbiamo già visto in fig:3.6 con λ di P5ter troviamo due parametri identici nella definizione del settore ϕ_R . Quindi oltre essere la strain che ha i dati maggiormente coerenti tra varie repliche

e FOV e anche quella che ci fornisce un proxy per il rate di crescita maggiormente coerente al framework che stiamo utilizzando. Si potrebbe pensare, in accordo con la letteratura, che la bontà di P5ter per quanto riguarda il growth rate è dovuta al fatto che è il promoter meno espresso (essendo un *-ter*) e quindi potrebbe essere meno influenzato da effetti che disturbino la crescita.

4 I nostri dati nel transiente di nutrienti falsificano l'ipotesi di quasi stazionarietà del modello di regolazione controllata dai flussi (FCR)

4.1 I dati sono analizzabili con il framework di Erickson e rispettano le leggi di crescita negli stati stazionari

A questo punto, definiti i proxy delle varie grandezze in gioco, possiamo analizzare i nostri dati attraverso il framework di Erickson[3] per osservare se vi sia coerenza tra i nostri e i suoi risultati.

La prima considerazione da fare è che l'ipotesi per cui $\chi_R(t) = \hat{\chi}_R(\sigma(t))$ (Fig.1.5) è l'ipotesi che vogliamo discutere e approfondire poichè i nostri dati mostrano che non è corretta, cosa che verrà ampiamente discussa in seguito, quindi non possiamo utilizzare l'equazione (1.6) che ne deriva direttamente. Perciò utilizziamo un set di equazioni, sempre prese dal framework di Erickson, che però sono definite senza bisogno che l'ipotesi sopracitata sia verificata:

$$\begin{cases} \phi_R^* = \phi_0 + \frac{\lambda^*}{\gamma} \\ \phi_{max} = \phi_P + \phi_R \\ \lambda_{bs}(t) = \sigma_A(t)\phi_R(t) \\ \frac{d\phi_R}{dt} = \sigma_A(t)\phi_R(t)[\chi_R(t) - \phi_R(t)] \end{cases} \quad (4.1)$$

La prima equazione è la growth law già vista in precedenza, da cui ricaviamo ϕ_R come spiegato nella sez.3.2 e dalla seconda ricaviamo $\phi_P = \phi_{max} - \phi_R$. Dopodiché possiamo calcolare σ_A utilizzando la terza equazione che sfrutta ϕ_R trovato prima e λ_{bs} che abbiamo come dato; infine possiamo inserire tutti questi ingredienti nell'ultima equazione per trovare χ_R (la derivata di ϕ_R è calcolata secondo quanto spiegato in append.A).

Iniziamo con l'osservare i dati negli stati stazionari, per vedere se si trovano sulle curve individuate da Erickson ([3], Extended Data Figure 1). In particolare ci interessa vedere se sono coerenti gli scatterplot di ϕ_R vs λ , sintomo di un'attinenza

dei nostri dati alla growth law (1.2) e χ_R vs σ che confermerebbe l'attinenza dei nostri dati al framework che stiamo utilizzando.

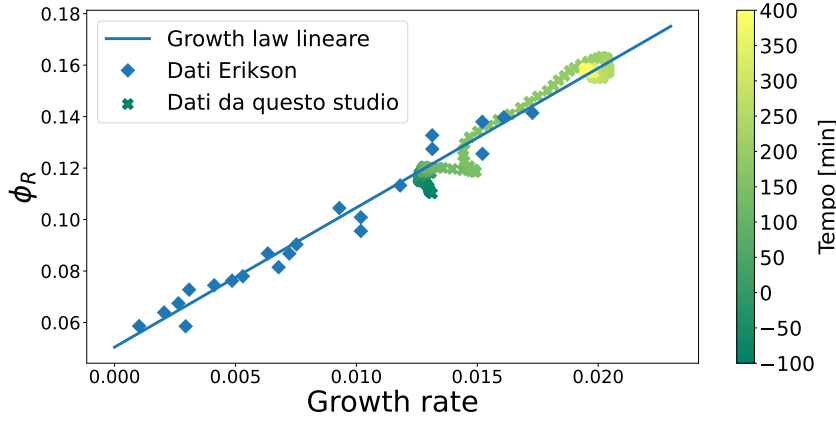


Figure 4.1: I dati di Erickson (estratti come spiegato in append.A) individuano una relazione lineare tra il settore ϕ_R e il rate di crescita già esplicitata come growth law, con un coefficiente angolare $\frac{1}{\gamma} \simeq 5,44[min]$ e un'intercetta $\phi_0 \simeq 0.049$. I due stati stazionari dei nostri dati, individuabili come quelle zone agli estremi dello scatterplot in cui abbiamo una maggior concentrazione di punti, si trovano esattamente sulla retta individuata dai parametri di Erickson. Abbiamo quindi una riconferma del fatto che i nostri dati seguano la growth law.

I due **steady state** si posizionano esattamente sulla retta individuata dalla growth law (Fig. 4.1). Si potrebbe obiettare il fatto che per definire ϕ_R abbiamo utilizzato la growth law stessa, quindi rischierebbe di non essere un risultato valido. Ricordiamo che, però, la retta nel grafico individuata dalla growth law (1.2) è rappresentata con i parametri γ e ϕ_0 (coefficiente angolare e intercetta) estratti da Erickson, quindi se i nostri stati stazionari giacciono su tale retta significa che la proporzionalità tra il nostro ϕ_R e il nostro λ segue la legge con tali parametri. Queste considerazioni vanno a rafforzare l'assunzione di bontà dei proxy scelti, difatti sapevamo già che le costanti $B_{1,2}$ che individuano ϕ_R dalla growth law (3.3) erano pressoché identiche (Fig.3.6), primo sintomo di coerenza. Ora sappiamo anche che, oltre a essere uguali, sono anche concordi con i parametri individuati da Erickson.

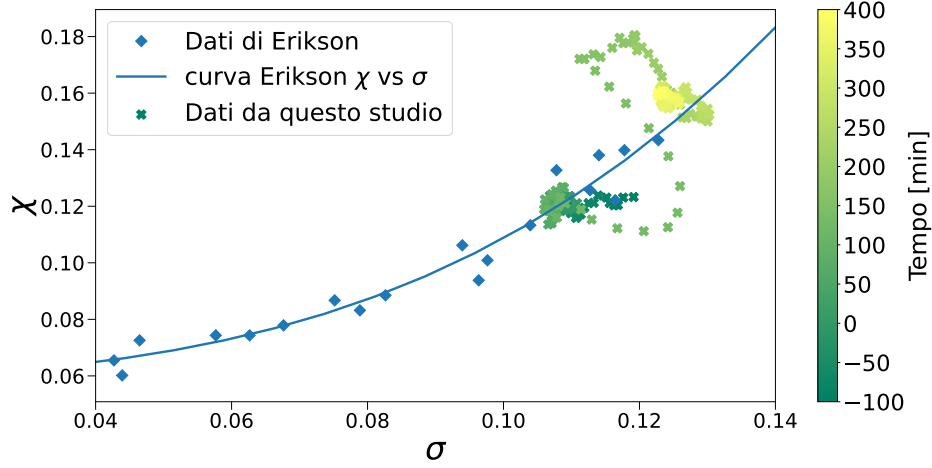


Figure 4.2: Lo scatterplot dei dati a steady state di χ_R vs σ individua una curva secondo Erickson. Si osserva chiaramente che gli steady state di χ_R vs σ ricavati dai nostri dati, attraverso le equazioni del framework (4.1), giacciono su tale curva.

Il plot 4.2 rappresenta forse uno dei risultati più importanti della nostra analisi, difatti è molto chiaro come gli **steady state** dei nostri dati si trovino precisamente sulla curva individuata dagli steady state di Erickson. Questo significa che, in primis, abbiamo un'ulteriore conferma della coerenza tra i nostri dati e il framework che stiamo utilizzando. Inoltre ci rivela che se limitiamo l'analisi ai soli stati stazionari effettivamente troviamo una relazione tra χ_R e σ , individuata dalla curva rappresentata nel grafico.

A questo punto però vi è da chiedersi se il fatto che esista una relazione tra i valori a steady state di χ_R e σ

$$\chi_R^* = \chi_R^*(\sigma^*)$$

possa tradursi, come afferma Erickson, nella già citata ipotesi

$$\chi_R(t) = \chi_R(\sigma(t)) \quad (4.2)$$

per cui χ_R sia una funzione di σ per qualsiasi tempo, compreso il periodo nel quale avviene lo shift.

4.2 I nostri dati falsificano l'ipotesi di quasi stazionarietà del modello FCR

Siamo arrivati al cuore di questa discussione: cercare di capire se l'ipotesi (4.2) sia valida. Possiamo già vedere dalla Fig. 4.2 come i punti che 'collegano' i due steady

state, che rappresentano i dati durante lo shift, escano totalmente dalla traiettoria delineata da Erickson e siano tali per cui è impossibile che $\chi_R = \chi_R(\sigma(t))$, per la definizione stessa di funzione. Cerchiamo quindi di approfondire questa analisi:

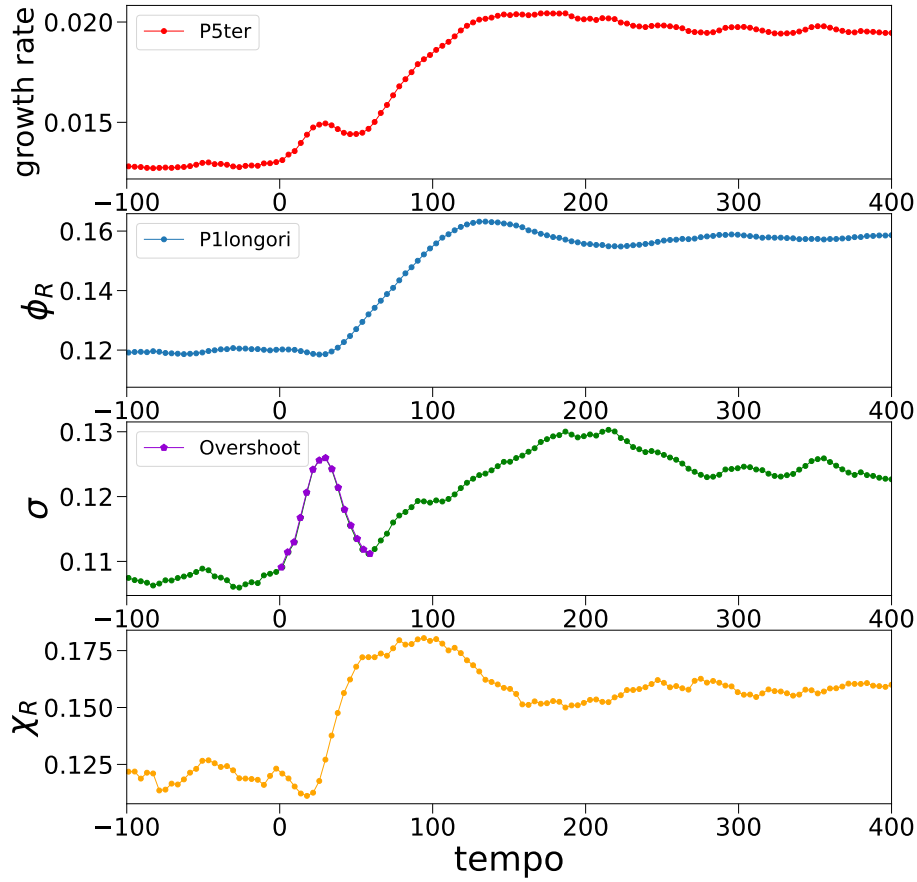


Figure 4.3: Sono riportate le serie temporali delle varie grandezze che compongono il nostro problema. Come già ribadito la serie di ϕ_R viene da P1longori, la serie del growth rate viene da P5ter. È stato marcato nel grafico di σ l'overshoot che si verifica subito dopo lo shift, per vedere successivamente che effetto ha questo overshoot nello scatterplot χ_R vs σ .

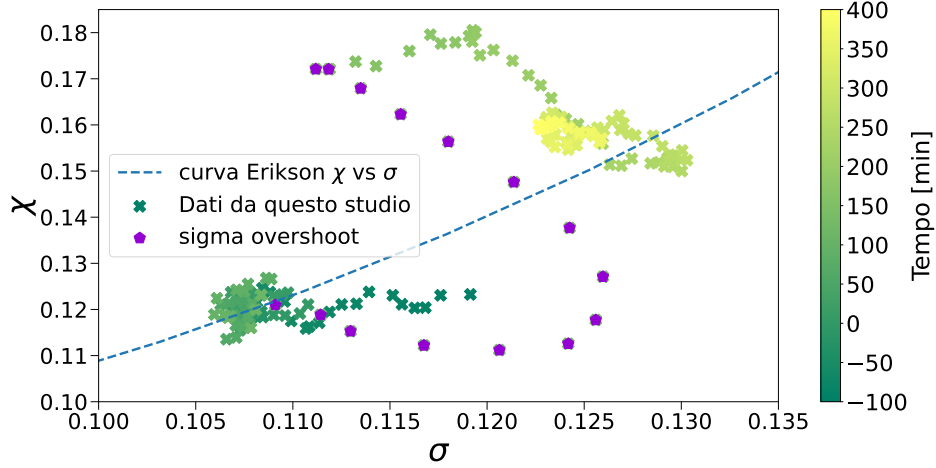


Figure 4.4: Come possiamo osservare dallo scatterplot, i punti corrispondenti all'overshoot di σ deviano completamente dalla curva dei dati di Erickson. Inoltre assumono una forma che rende impossibile definire $\chi_R(t) = \chi_R(\sigma(t))$ poichè per lo stesso valore di σ abbiamo due valori di χ_R per tutto il tempo dell'overshoot.

Questo primo test evidenzia come l'ipotesi (4.2) non regga per i nostri dati a causa dell'overshoot di σ che anticipa la crescita di χ_R (Fig. 4.3 e 4.4). Per rafforzare questa tesi abbiamo provato a 'tagliare' l'overshoot con delle semplici interpolazioni lineari (vedi append.A), osservando il risultato nello scatterplot.

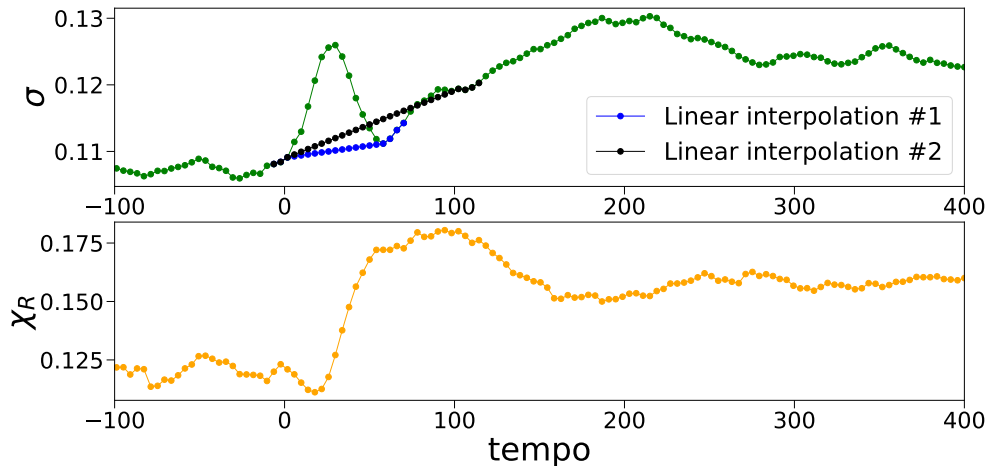


Figure 4.5: In questo caso, partendo dallo stesso grafico visto in precedenza, facciamo delle interpolazioni lineari semplici che 'tagliano' l'overshoot di σ , cosicchè cresca più lentamente e in modo più regolare, per testare come la crescita di χ influenzi lo scatterplot.

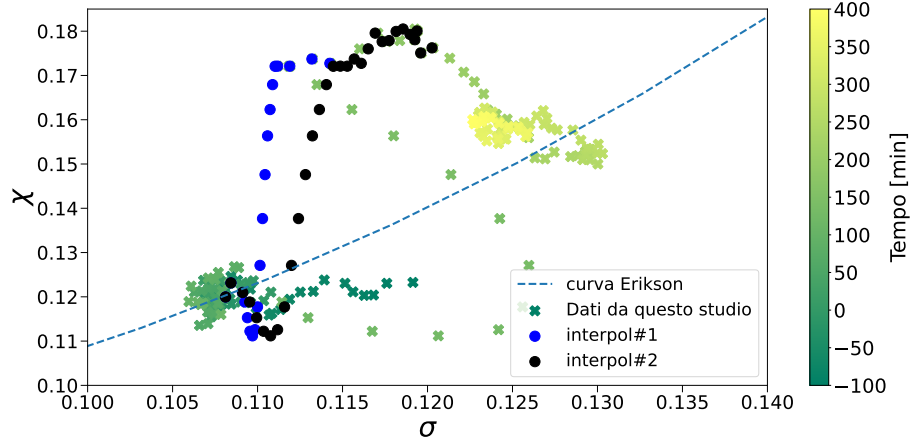


Figure 4.6: Vediamo come, pur eliminando l'overshoot di sigma, otteniamo una figura nello scatterplot lontana dalla curva prevista da Erickson. Difatti in questo caso, in cui è stata eliminata la crescita piccata e repentina di σ osserviamo due curve quasi 'a gradino' che collegano i due steady state, dovute alla crescita importante di χ_R

Siamo arrivati ad affermare che durante uno shift di nutrimenti, non si possa ritenere vera l'ipotesi di quasi-staticità avanzata nel modello FCR per cui il sistema segua la curva $\chi_R = \chi_R(\sigma(t))$, in una serie di stati metastabili per tutti i tempi tra uno steady state e l'altro. Possiamo quindi dire che l'ipotesi (4.2) di quasi-staticità del modello FCR è un'approssimazione (come definita dagli stessi autori) temporale "a grana grossa", che non tiene conto delle diverse scale temporali su cui si muovono σ e χ_R durante lo shift, portando il sistema fuori da un regime di quasi-stazionarietà.

Modelli

5 Un modello minimale in cui σ e χ_R sono assegnati indipendentemente riproduce qualitativamente i dati nel transiente

Cerchiamo di sfruttare, a questo punto, le equazioni del framework (4.1) che non implicano l'ipotesi (4.2), rivelatasi inconsistente, per definire un modello che riproduca i dati sperimentali. Il primo tentativo consiste in un modello in cui assegniamo manualmente le funzioni di σ e χ_R , assumendone una forma 'a gradini'. Seppur riconoscendo che gli input del modello vengono messi manualmente riusciamo a catturare tutte le caratteristiche qualitative dei dati, rispettando le loro diverse scale temporali. Difatti vediamo chiaramente in Fig. 5.1 che la

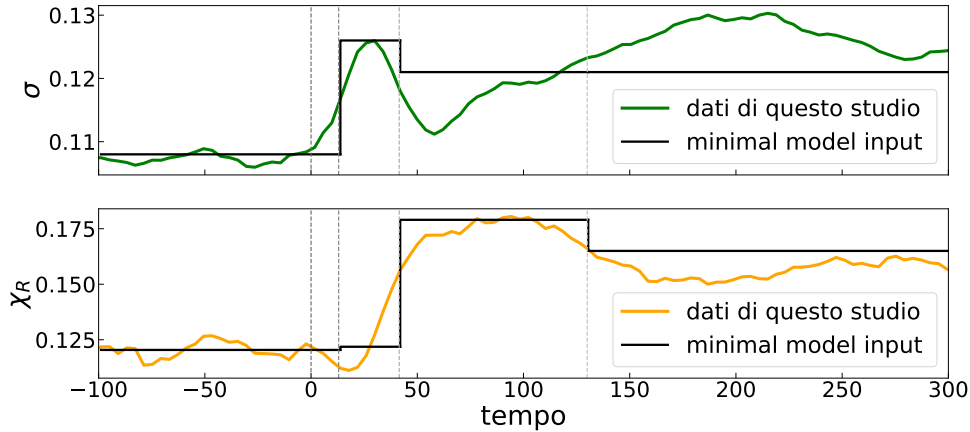


Figure 5.1: Sono rappresentati in nero i valori "a gradini" di σ e χ_R utilizzati come input del modello minimale descritto. Si può vedere chiaramente come l'overshoot di σ preceda la crescita di χ_R , si abbassi nel momento in cui χ_R si alza verso il suo valore massimo, più alto dello stato stazionario a cui si stabilizza successivamente.

crescita di χ_R al suo valore massimo è ritardata rispetto all'overshoot di σ , che avviene pressoché istantaneamente dopo lo shift di nutrienti. La funzione

regolatrice dei ribosomi χ_R ha invece bisogno di più tempo per adattarsi alle nuove condizioni, inizia infatti a crescere verso il suo valore massimo quando l'overshoot di σ comincia a scendere, per poi assestarsi infine sul valore ϕ_{Rf}^* .⁴ Possiamo quindi riassumere l'input del modello minimale con:

$$\sigma_A(t) = \begin{cases} \sigma_i^* & t < t_{start} \\ \sigma_{max}^* & t_{start} < t < t_\sigma \\ \sigma_f^* & t > t_\sigma \end{cases} \quad \chi_R(t) = \begin{cases} \chi_{Ri}^* & t < t_\sigma \\ \chi_{Rmax}^* & t_\sigma < t < t_\chi \\ \phi_{Rf}^* & t > t_\chi \end{cases}$$

dove $t_{start} \simeq 15min$, $t_\sigma \simeq 45min$, $t_\chi \simeq 135min$

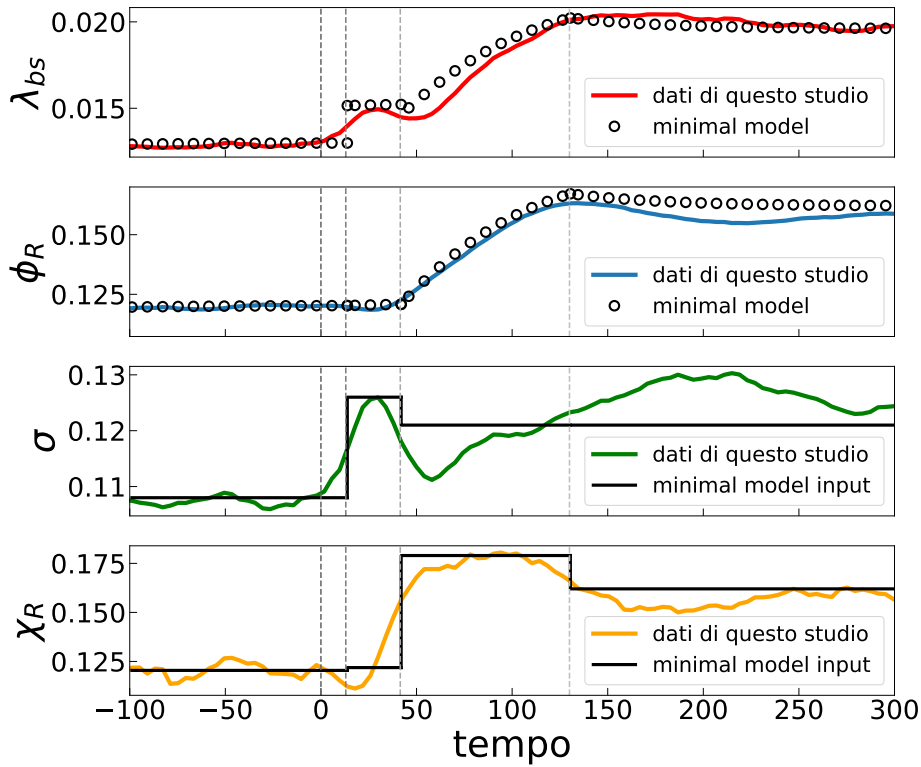


Figure 5.2: Il nostro modello minimale riesce a rappresentare tutte le caratteristiche qualitative dei dati

Siamo riusciti a trovare quindi, all'interno del framework, un modello minimale che rappresenti qualitativamente le scale temporali e gli overshoot presenti

⁴Il valore finale che raggiunge χ_R dopo lo shift è pari al valore finale che raggiunge ϕ_R ; come possiamo vedere dall'equazione 4 di (4.1) devono essere uguali affinché si abbia uno stato stazionario di ϕ_R .

nei dati (Fig. 5.2). Il limite di questo modello è rappresentato dalle funzioni di input che vengono assegnate manualmente, mentre vorremmo trovare un modello meccanico che funzioni ugualmente.

6 Un modello meccanicistico dell'allocazione dei ribosomi nel transiente devia dall'assunzione di quasi-stazionarietà, ma non riproduce gli overshoot nei dati

Dato che il modello fenomenologico fornisce una comprensione limitata delle cause delle deviazioni dei dati da quanto atteso dalle teorie standard, il passo successivo che abbiamo effettuato è stato quello di cercare un modello meccanicistico che riuscisse a descrivere i nostri dati. Il primo tentativo esplorato cerca di includere gli effetti della trascrizione dei ribosomi attraverso la regolazione del ppGpp, che setta χ_R e introduce una ulteriore scala temporale nel sistema. L'osservazione dell'oscillazione di χ durante il transiente poteva suggerire una dinamica tipo oscillatore armonico smorzato, in regime di smorzamento critico. Idealmente un buon modello dovrebbe fornire alla un'equazione differenziale per χ che ammetta soluzioni di questo tipo, per riprodurre i dati. Bisogna sottolineare che in questo caso, a differenza del modello minimale, stiamo cercando un modello indipendente per χ , che non coinvolga le altre grandezze, per poi utilizzare le equazioni del framework (4.1) per ricavare σ , utilizzando come input il dato di ϕ_R .

Per definire questo modello dobbiamo introdurre ω come l'occupazione dell'RNAP (RNA Polimerasi) di siti R o P ovvero la probabilità che RNAP stia leggendo RNA ribosomale o costitutivo, settata dai livelli di ppGpp. Nel formalismo dei settori avremo naturalmente $\omega_R + \omega_P = \omega_{max}$ con $\omega_{max} = \phi_{max} \simeq 0.55$. Assumiamo che i livelli di ppGpp varino quasi istantaneamente[2] e che il rilassamento di ω durante uno shift sia associato ad un'unica scala temporale, secondo un'equazione del tipo

$$\frac{d\omega_R}{dt} = -\tau_\omega^{-1} (\omega_R - \bar{\omega}_R) , \quad (6.1)$$

dove τ^{-1} è il tempo tipico di questo rilassamento e il target è per definizione $\bar{\omega}_R = \frac{1}{1 + \frac{[G]_{free}}{k_g}} \omega_0$ dove $\frac{[G]_{free}}{k_g}$ è una costante legata alla ppGpp e ω_0 è l'occupazione a $[G] = 0$. Siamo in grado di individuare nei nostri dati questo target poiché risulta essere $\bar{\omega}_R = \phi_{Rf}^*$. A questo punto definiamo due equazioni, una per la massa dei trascritti totali T^5 e per la massa dei trascritti ribosomali T_R presenti

⁵per trascritti si intendono gli RNA messaggeri, mRNA, ovvero il prodotto della trascrizione

nella cellula:

$$\frac{dT}{dt} = \beta P - \delta T \quad (6.2)$$

$$\frac{dT_R}{dt} = \beta \omega_R P - \delta T_R, \quad (6.3)$$

dove β è una costante di massa, P è il numero di RNAP, e il termine negativo è un termine di degradazione tramite la costante δ che descrive il processo di eliminazione dei trascritti nella cellula. L'assunzione principale che facciamo è definire

$$\chi_R = \frac{T_R}{T}$$

dopodiché svolgendo $\frac{d\chi_R}{dt} = \frac{d}{dt} \left(\frac{T_R}{T} \right)$ otteniamo:

$$\frac{d\chi_R}{dt} = (\chi_R - \omega_R) \quad (6.4)$$

da cui, inserendo (6.1), si ottiene un'equazione differenziale del secondo ordine per la funzione di regolazione,

$$\frac{d^2\chi_R}{dt^2} + \left(\frac{1}{\tau_\omega} + \frac{1}{\tau_\chi} \right) \frac{d\chi_R}{dt} + \frac{1}{\tau_\chi \tau_\omega} (\chi_R - \bar{\omega}_R) = 0, \quad (6.5)$$

con $\frac{1}{\tau_\chi} = \frac{\beta P}{T}$. Conoscendo χ_R e potendo quindi ricavarne le derivate, non ci rimane che trovare i due parametri liberi τ_ω e τ_χ per decretare se il modello possa riprodurre i dati. L'algoritmo di fit⁶ non riesce a restituirci dei valori validi per i due parametri perché effettivamente non esistono tali da avere un'oscillazione smorzata come quella che stiamo cercando. Pur provando a variare manualmente i parametri non riusciamo mai a trovare una forma che rappresenti i dati.

della RNAP. In breve: l'RNAP legge il codice genetico e lo trascrive in "istruzioni", rappresentate dai mRNA, che serviranno ai ribosomi come ricetta per tradurre le proteine. Naturalmente nel formalismo dei settori vengono definiti i trascritti ribosomali e costitutivi, rispettivamente come mRNA che portano le istruzioni per la produzione di ribosomi o proteine costitutive

⁶scipy.optimize.curve_fit

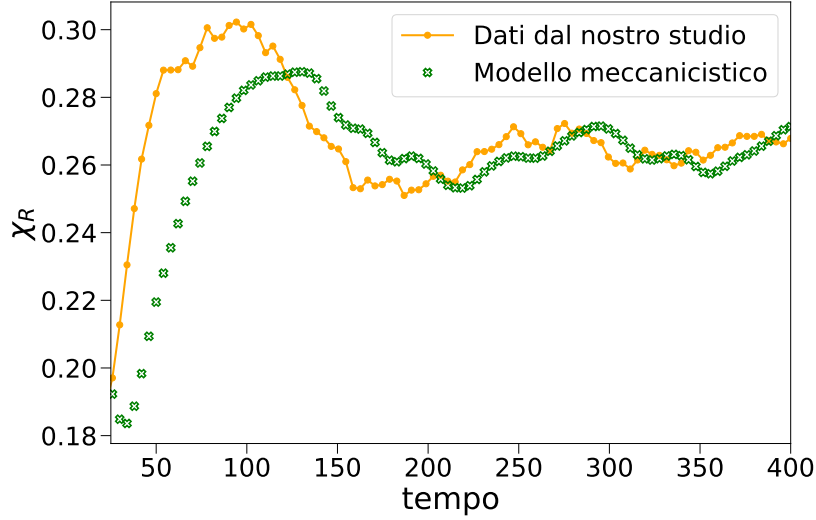


Figure 6.1: Sebbene sia possibile trovare dei parametri che rappresentino un' oscillazione, il modello non riesce a rappresentare correttamente i dati sperimentali.

7 Proposta di un modello con una equazione per la concentrazione di amminoacidi, con σ legato al suo valore

L'ultima proposta di modello nasce sempre dalla necessità di spiegare il rapido overshoot di σ che porta la dinamica di crescita nel transiente fuori dalla quasi-stazionarietà. Ricordiamo che nel nostro apparato sperimentale lo shift di nutrienti è rappresentato da un flusso di amminoacidi che viene fornito alla coltura di *E.coli*. I nuovi amminoacidi sono quindi immediatamente disponibili per la sintesi proteica senza dover passare dal processo di demolizione come i nutrienti derivanti da sorgenti di carbonio.

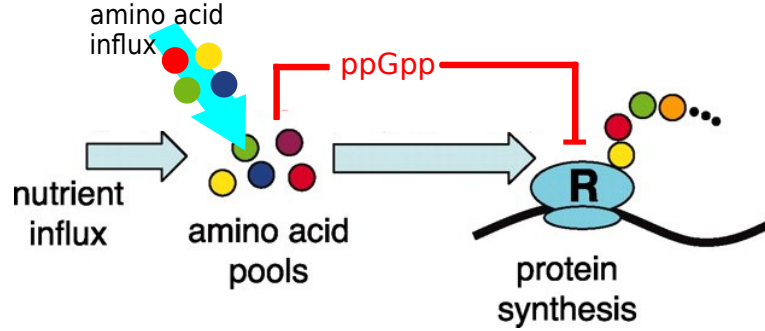


Figure 7.1: Illustrazione semplificata del flusso diretto di aminoacidi.

Vogliamo capire se questo possa essere il motivo per cui osserviamo un'overshoot nel rate di traduzione σ che è sicuramente dipendente dal numero di nutrienti disponibili. Scriviamo le equazioni di questo modello riprendendo il formalismo FCR[3] dei flussi, e confrontandole con uno dei modelli [4] esistenti che cerca di descrivere la dinamica cellulare durante lo shift. Nel primo stato stazionario abbiamo il flusso di nutrienti 'in entrata' J_{in} determinato dalla sorgente di carbonio iniziale bilanciato con il flusso di traduzione delle proteine J_R , che conosciamo dalla definizione (1.4). Trovandoci in uno steady state dobbiamo necessariamente avere un bilancio pari dei flussi: tutti i nutrienti che vengono assorbiti dalla cellula devono essere consumati, altrimenti avremmo un aumento (o diminuzione) del rate di crescita, in caso di eccesso (o difetto) di nutrienti assorbiti rispetto a quelli consumati.

$$J_{in,1}^* = kM\phi_{P1}^* \quad (7.1)$$

$$kM\phi_{P1}^* = J_R = \lambda_1^*M \quad (7.2)$$

dove M è la massa totale. Dall'equazione di bilancio (7.2) ricavo la costante $k = \frac{\lambda_1^*}{\phi_{P1}^*}$. Al tempo $t = 0$ aggiungiamo gli aminoacidi che possiamo formalizzare come un flusso J_A , ottenendo le seguenti equazioni per lo stato stazionario finale:

$$J_{in,2}^* = kM\phi_{P2}^* + J_A \quad (7.3)$$

$$kM\phi_{P2}^* + J_A = J_R = \lambda_2^*M \quad (7.4)$$

Da (7.4) inserendo al costante k , ricavo

$$J_A = \lambda_2^* - \lambda_1^* \frac{\phi_{P2}^*}{\phi_{P1}^*}$$

A questo punto sono in grado di scrivere l'equazione che racchiude la dinamica del transiente, partendo sempre dall'idea che il numero di nutrienti sia il fattore

preponderante da cui dipende il rate di traduzione di proteine

$$\frac{dA}{dt} = J_{in} - J_R \quad (7.5)$$

$$\frac{dA}{dt} = \frac{\phi_{P2}^*}{\phi_{P1}^*} M \phi_P + J_A - \lambda M \quad (7.6)$$

C'è bisogno ancora di perfezionare questo modello poiché ci sono dei parametri liberi da determinare, va determinata la relazione (o quantomeno la costante di proporzionalità) tra $A(t)$ e $\sigma(t)$, sempre se la dinamica di A riproduca correttamente quella di σ . Difatti penso che questo modello non sia ancora in grado di descrivere i nostri dati, poiché per quanto possa essere presente un overshoot nella dinamica di A descritta in (7.6), il picco di sigma nei dati si consuma molto velocemente, non ha una forma di una curva che "rilassa" come potremmo trovare da (7.6). È possibile che ci siano altri fattori da dover considerare ad esempio l'attivazione istantanea di ribosomi inattivi.

Conclusioni

Questa tesi concerne le leggi che regolano la crescita cellulare nel formalismo dei settori. A partire da un data set di upshift di nutrienti prodotto dai nostri collaboratori di Cambridge tramite esperimenti di crescita di *E. coli* in chip microfluidici high-throughput[7], il nostro scopo è stato produrre modelli quantitativi di allocazione delle risorse nella crescita cellulare che potessero descrivere i dati.

La prima parte (sezione 2) definisce le osservabili macroscopiche. Dopo varie considerazioni, abbiamo definito $P_{longori}$ come miglior proxy per il settore ribosomale e P_{5ter} come proxy per il rate di crescita. Ancor più importante, abbiamo mostrato la coerenza dei nostri dati a steady state con le growth law note. La scelta di questi proxies è giustificata dai grafici suddivisi per FOV in cui sono evidenziati degli errori sistematici in alcuni dati, che vengono quindi scartati.

Il capitolo 4.2 ha mostrato come una delle ipotesi fondamentali del modello FCR (4.2), quella del bilancio istantaneo dei flussi[3], non sia coerente con i nostri dati durante lo shift di nutrienti. Questa conclusione è supportata dal fatto che a steady state i nostri dati sono perfettamente coerenti con le predizioni di questo modello, ma durante lo shift deviano dalla condizione di quasi-staticità ipotizzata da Erikson nel transiente. Inoltre abbiamo constatato che la funzione di regolazione χ_R non può essere espresso come funzione del rate di traduzione per ribosoma σ . Analizzando i punti in cui questa relazione devia dalla quasi-stazionarietà durante il transiente, abbiamo individuato due fattori come causa della falsificazione dell'ipotesi (4.2). Per prima cosa si ha un'overshoot di σ che avviene in anticipo rispetto alla crescita di χ , probabilmente dovuto all'eccesso di ribosomi[4, 5] e/o all'accumulazione di amino acidi immediatamente dopo lo shift. Inoltre, pur non considerando il picco di σ , abbiamo mostrato che anche l'overshoot di χ_R devia la relazione dalla quasi-stazionarietà su tempi più lunghi dopo lo shift.

Queste considerazioni mostrano l'impossibilità di descrivere i dati attraverso una singola scala temporale, portandoci a proporre diversi modelli (capitolo III) che possano rispettare correttamente queste caratteristiche dei dati. Il primo modello rappresenta bene i dati e le loro scale temporali, ma è puramente fenomeno-

logico, e richiede input manuali di σ e χ , inseriti come step function su scale temporali ad hoc. Abbiamo cercato poi di definire modelli più microscopici/meccanicistici, che includessero (i) la regolazione trascrizionale tramite ppGpp, che porta a una equazione di evoluzione di χ_R come oscillatore armonico smorzato e (ii) l'accumulazione dinamica degli amino acidi nella cellula, sfruttando il formalismo dei flussi, che trova molte analogie con il modello generico proposto di Kohanim *et. al* [4] presente in letteratura.

A Metodi

A.1 Serie temporali

Fluorescenza/Volume

Le serie temporali di Fluorescenza/Volume sono state ottenute con una media binnata sul tempo, con punti spaziati ogni 4 minuti e una finestra temporale su cui viene calcolata la media di 20 min (± 10 min).

Growth Rate

Il growth rate di singola cellula è stato calcolato applicando un filtro di Savitzky-Golay ⁷ sui volumi misurati ogni 5 minuti della cellula e dopodiché calcolando il growth rate come

$$\lambda_{bs}(t_i) = \frac{1}{v(t_i)} \frac{v(t_{i+1}) - v(t_{i-1})}{t_{i+1} - t_{i-1}}$$

A.2 Estrazione dati Erickson *et al.*

I dati di Erickson *et al.* sono stati estratti direttamente dai plot presenti in [3] attraverso un software online capace di estrarre i dati da un plot punto per punto.⁸ Le curve che fittano i dati di Erickson sono state individuate rispettivamente con un fit lineare (fig.4.1) e con un fit polinomiale di grado 3 (fig.4.2) con la funzione *numpy.polyfit*.

A.3 Interpolazioni

Le interpolazioni nella sezione 4.2 sono state fatte con *scipy.interpolate.interp1d* con il metodo *'linear'*, eliminando i dati dell'overshoot di σ e facendoli sostituire da quelli che venivano calcolati da tale algoritmo.

⁷un filtro statistico che compie un operazione di smoothing dei dati, eliminando un po' del rumore in eccesso che potrebbe dare molto fastidio nel calcolo di una derivata "discreta", come quello che facciamo in questo caso

⁸<https://automeris.io/WebPlotDigitizer/>

B Confronto tra il nostro modello sulla concentrazione di aminoacidi e quello di Kohanim *et al.*

La dinamica pensata nel modello di Kohanim *et al.* è simile a quella del nostro modello. Non viene utilizzato esplicitamente il concetto di settori ma viene riconosciuta la suddivisione del proteoma in proteine ribosomali e costitutive, con una funzione di ripartizione (la nostra χ_R) che gestisce questa dinamica. Introduce due funzioni, $h(x)$ e $g(x)$, che rappresentano rispettivamente il grado di saturazione (hanno valori da 0 a 1) del processo di trascrizione (h) e traduzione (g), dove x è il numero di nutrienti disponibili.

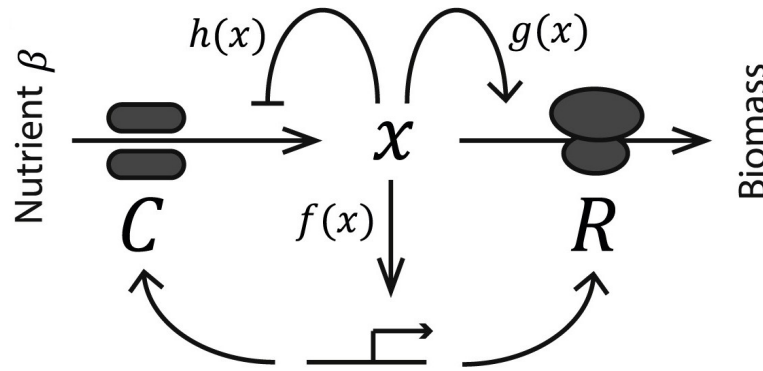


Figure B.1: Illustrazione semplificata del modello di Kohanim. Il processo illustrato è esattamente l'oggetto di questa tesi, vengono solo modellizzate in modo un po' diverso le funzioni h e g che nel nostro modello sono determinate dai settori. La funzione f rappresenta la nostra χ_R , funzione che descrive la ripartizione del proteoma dettata dalla ppGpp.

Cerchiamo ora di osservare in parallelo tutte le grandezze e le equazioni rispetto al nostro modello: viene definito il rate di crescita come

$$\mu = \mu_{sat}g(x)R(x)$$

μ_{sat} rappresenta il rate di crescita del secondo stato stazionario e $R(x)$ la quantità di ribosomi. Da questo troviamo che il nostro rate di traduzione $\sigma \sim \mu_{sat}g(x)$. Dopodiché viene definito un set di equazioni che descrivano la dinamica in funzione dei nutrienti:

$$\frac{dR}{dt} = \mu(f(x) - R) \tag{B.1}$$

$$\frac{dx}{dt} = \beta C h(x) - \mu \tag{B.2}$$

B CONFRONTO TRA IL NOSTRO MODELLO SULLA CONCENTRAZIONE DI AMINOACIDI E QUELLO DI KOHANIM *ET AL*41

la prima delle quali è identica all'ultima delle equazioni (4.1), dove $f(x)$ è il nostro χ_R . La seconda presenta un termine β che è definito 'disponibilità di nutrienti', tipo il nostro k dell'equazione (7.2); il termine C rappresenta il numero di proteine costitutive e $h(x)$ la saturazione del processo di demolizione del carbonio in aminoacidi. Difatti questa equazione è uguale a (7.6) senza il termine di flusso di aminoacidi che abbiamo aggiunto nel nostro caso.

Il punto cruciale di questo modello è che una volta avvenuto lo shift si passa istantaneamente da una saturazione di traduzione $g(x_0)$ alla saturazione massima $g(x_1) = 1$, riconoscendo che la funzione di ripartizione $f(x)$ risponde in ritardo rispetto al cambiamento di nutrienti. Si ha quindi:

$$\begin{aligned}\mu_{pre-shift} &= \mu_{max} f(x_0) g(x_0) \\ \mu_{post-shift} &= \mu_{max} f(x_0) g(x_1)\end{aligned}$$

Abbiamo quindi un modello che si colloca 'a metà' tra la nostra idea e il modello FCR. Difatti viene riconosciuto che durante lo shift non si ha un bilancio tra il flusso di nutrienti in entrata e il flusso di traduzione di proteine, come ipotizza invece il modello FCR, poiché si avrebbe una variazione nulla della disponibilità di nutrienti che non permetterebbe al sistema di evolversi verso un nuovo stato stazionario con un rate di crescita maggiore, secondo (B.2). Afferma inoltre che la funzione di risposta $f(x)$ reagisca in ritardo rispetto alla variazione di nutrienti, discostandosi dalla ipotesi di quasi stazionarietà del modello FCR. Tuttavia condivide con il modello FCR l'idea per cui σ (nel loro caso è $g(x)$) e quindi λ (nel loro caso μ) evolvano in una dinamica 'a salto' passando quasi istantaneamente dal valore iniziale al valore finale.

References

- [1] Evert Bosdriesz, Douwe Molenaar, Bas Teusink, and Frank J Bruggeman. How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization. *The FEBS journal*, 282(10):2029–2044, 2015.
- [2] Ferhat Büke, Jacopo Grilli, Marco Cosentino Lagomarsino, Gregory Bokinsky, and Sander Tans. ppgpp is a bacterial cell size regulator. *bioRxiv*, 2020.
- [3] David W Erickson, Severin J Schink, Vadim Patsalo, James R Williamson, Ulrich Gerland, and Terence Hwa. A global resource allocation strategy governs growth transition kinetics of escherichia coli. *Nature*, 551(7678):119–123, 2017.
- [4] Yael Korem Kohanim, Dikla Levi, Ghil Jona, Benjamin D Towbin, Anat Bren, and Uri Alon. A bacterial growth law out of steady state. *Cell reports*, 23(10):2891–2900, 2018.
- [5] Jennifer L McGinnis, Qi Liu, Christopher A Lavender, Aishwarya Devaraj, Sean P McClory, Kurt Fredrick, and Kevin M Weeks. In-cell shape reveals that free 30s ribosome subunits are in the inactive state. *Proceedings of the National Academy of Sciences*, 112(8):2425–2430, 2015.
- [6] Matteo Osella, Eileen Nugent, and Marco Cosentino Lagomarsino. Concerted control of escherichia coli cell division. *Proceedings of the National Academy of Sciences*, 111(9):3431–3435, 2014.
- [7] Mia Panlilio, Jacopo Grilli, Giorgio Tallarico, Bianca Sclavi, Pietro Cicuta, and Marco Cosentino Lagomarsino. Threshold accumulation of a constitutive protein explains e. coli cell division behavior in nutrient upshifts. 2020.
- [8] Matthew Scott, Carl W Gunderson, Eduard M Mateescu, Zhongge Zhang, and Terence Hwa. Interdependence of cell growth and gene expression: origins and consequences. *Science*, 330(6007):1099–1102, 2010.
- [9] Irit Shachrai, Alon Zaslaver, Uri Alon, and Erez Dekel. Cost of unneeded proteins in e. coli is reduced after several generations in exponential growth. *Molecular cell*, 38(5):758–767, 2010.
- [10] Ping Wang, Lydia Robert, James Pelletier, Wei Lien Dang, Francois Taddei, Andrew Wright, and Suckjoon Jun. Robust growth of escherichia coli. *Current biology*, 20(12):1099–1103, 2010.