

**ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA
CAMPUS DI CESENA
SCUOLA DI INGEGNERIA E ARCHITETTURA**

CORSO DI LAUREA MAGISTRALE IN INGEGNERIA BIOMEDICA

**MODELLAZIONE MATEMATICA DELLE
DINAMICHE DI RILEVAMENTO
DELL'ESPRESSIONE DEL GENE *N-MYC*
MEDIANTE UN SENSORE
BIOMOLECOLARE SINTETICO**

Elaborato in
Bioingegneria cellulare e molecolare

Relatore
Prof. Emanuele D. Giordano

Presentato da
Eleonora Manoni

Correlatrici
Ing. Lucia Bandiera
Ing. Marilisa Cortesi

Sessione III
Anno Accademico 2014/2015

INDICE

1. Introduzione.....	5
1.1 Neuroblastoma e MYCN.....	5
1.1.1 Il gene MYCN.....	5
1.1.2 Amplificazione vs. sovraespressione di MYCN.....	8
2. Scopo della tesi.....	9
3. Materiali e Metodi.....	11
3.1 Componenti molecolari del circuito (<i>biological parts</i>).....	11
3.1.1 Promotori.....	11
3.1.2 Terminatori.....	12
3.1.3 Siti di legame al ribosoma (RBS).....	13
3.1.4 Sequenze codificanti.....	13
3.1.5 Chassis.....	13
3.2 Standard assembly.....	14
3.3 Topologia e principio di funzionamento del circuito che implementa il sensore molecolare.....	16
3.4 Modellazione della dinamica circuitale.....	18
3.5 Reazioni ed equazioni del sistema.....	18
3.6 Parametri del modello.....	21
3.7 Identificazione dei ratei trascrizionali.....	23
4. Risultati e discussione.....	27
4.1 Identificazione dei tratti CIS e TRANS.....	27
4.2 Identificazione dei parametri.....	28
5. Conclusioni.....	35

6. Ringraziamenti.....	37
7. Bibliografia.....	49
8. Sitografia.....	41

1. Introduzione

1.1 Neuroblastoma e MYCN

Il *Neuroblastoma* (NBL) è il più comune tumore extra-cranico pediatrico, ed è associato a uno dei più alti tassi di mortalità oncologica infantile. Clinicamente si distinguono due forme di NBL: una localizzata e a decorso solitamente benigno, e una metastatica e ad alto rischio. Nel 25-30% dei casi lo sviluppo di NBL è direttamente associato all'attivazione (sovraespressione) del protooncogene *MYCN* [<https://ghr.nlm.nih.gov/gene/MYCN>], spesso tramite *amplificazione genica* - un processo che porta all'aumento del numero di copie di un gene a causa di mutazioni somatiche (non ereditarie ma dovute a errori della replicazione cellulare) e favorisce l'insorgenza di un fenotipo tumorale [Chayka 2015]. Per questo motivo i livelli di amplificazione genica nei pazienti sono attualmente adottati come riferimento prognostico della severità del NBL.

Tuttavia, l'amplificazione genica è solo uno dei possibili fattori di iperattività di *MYCN*, che può essere sovraespresso anche in assenza di una sua amplificazione aberrante.

1.1.1 Il gene MYCN

MYCN è situato sul braccio corto del cromosoma 2, dove è stato mappato nella regione 2p23-24 mediante ibridazione in situ [Nesbit 1999]. Normalmente la sua espressione è limitata alle prime fasi embrionali dove, durante lo sviluppo della cresta neurale, è coinvolto nella regolazione del fenotipo cellulare neurale [Purgato 2008].

Tuttavia, *MYCN* è un proto-oncogene. I proto-oncogeni sono geni normalmente coinvolti nel controllo della proliferazione cellulare e a causa di diversi fattori, come l'amplificazione genica o mutazioni, possono divenire oncogeni, ovvero indurre la trasformazione di cellule sane in cancerose (<http://ghr.nlm.nih.gov/gene/MYCN>).

La famiglia dei geni *MYC*, composta da *MYC* (o *c-Myc*), *MYCN* e *MYCL*, codifica per un gruppo di proteine nucleari che svolgono un ruolo nella proliferazione cellulare e nella regolazione del ciclo cellulare, nel differenziamento, nell'apoptosi, nell'instabilità genomica e nella trasformazione neoplastica. In particolare, la proteina codificata da *MYCN* sembra avere un ruolo di rilievo nella regolazione del differenziamento cellulare e nello sviluppo neuroectodermico [Shoji et al 2015]. L'oncoproteina *MYCN*, espressa in organismi adulti, regola la trascrizione di geni che potenzialmente possono promuovere la genesi del tumore.

MYCN è stato originariamente clonato nel 1983, identificando una sequenza di DNA amplificato con parziale omologia col proto-oncogene *MYC* in linee cellulari di NBL.

Il gene si compone di 3 esoni (I 1-183, II 184-1090, III 1091-2605): il primo è soggetto a splicing alternativo e gli altri codificano allo stesso modo per la proteina; la sua trascrizione inizia in due siti che sono sotto il controllo di due promotori alternativi. Questo aspetto, combinato allo *splicing* alternativo genera due forme di mRNA. I messaggeri hanno diverse sequenze leader 5', ma presentano corpi identici formati dal secondo e dal terzo esone. Entrambe le forme di mRNA hanno un'emivita breve (ca 15 minuti) e codificano per due diverse proteine [Stanton 1987].

N-myc è una fosfoproteina localizzata nella matrice nucleare, che ha un'emivita di 30-50 minuti e presenta regioni a elevata omologia con le altre proteine della famiglia.

MYCN è costituita da:

- Un dominio N-terminale di transizione;

- Una regione intermedia non strutturata;
- Una regione C-terminale contenente un dominio basico (BR), coinvolto nel legame sequenza-specifico con il DNA, e un motivo *helix-loop-helix zipper* (bHLH-LZ), che media il legame al DNA e l'interazione con altre proteine bHLH-LZ, come Max e Mad.

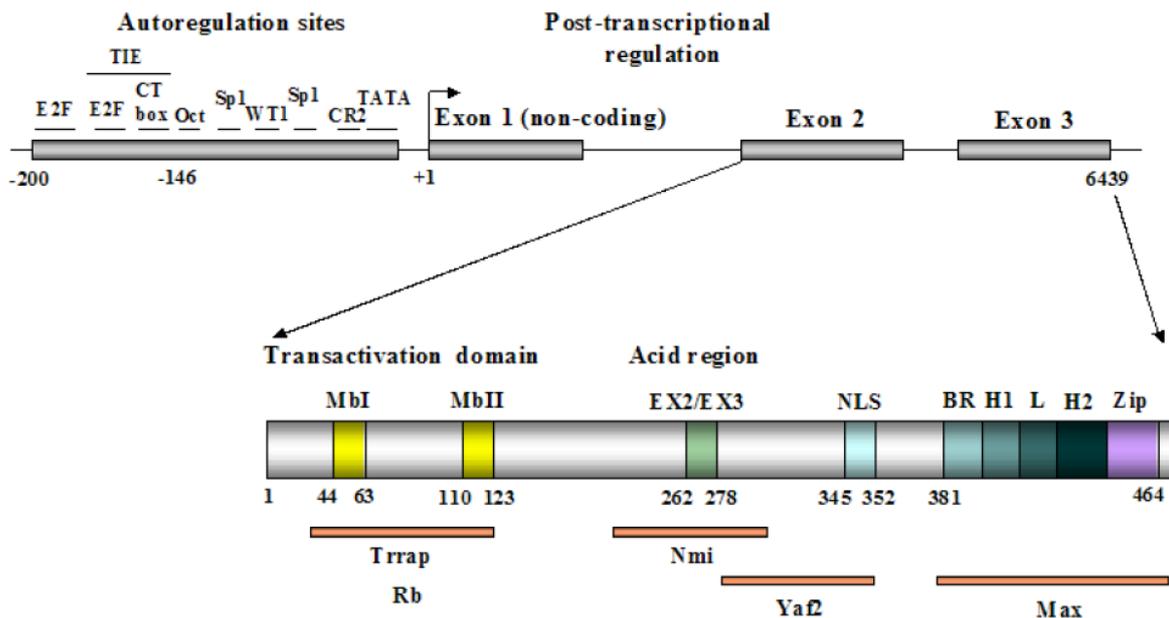


Figura 1: Il gene MYCN, dalla immagine è possibile visualizzare i tre diversi esoni e la sequenza codificante comune ad entrambe le proteine che derivano dal gene [Purgato 2008].

Per legarsi al DNA, N-MYC deve prima dimerizzare con Max. In seguito al legame col DNA, la proteina regola l'espressione genica e interviene nel controllo della proliferazione del differenziamento cellulare. Il dimero N-MYC/Max si lega al DNA in modo specifico alla sequenza CACGTG, denominata *E-box*. Un altro sito di legame è costituito dalla sequenza CATGTG. Il dominio transattivatore di N-MYC, posto all'N-terminale, è coinvolto nell'equilibrio tra segnali apoptotici e proliferativi e costituisce anche il sito di legame di proteine che regolano l'attività di N-MYC. L'attività del dimero N-MYC

/Max è regolata da un'altra proteina, Mad, che, formando eterodimeri con Max, risulta in competizione con MYC. Il dimero Myc/Max attiva numerosi geni, tra i quali:

- Il regolatore della proliferazione cellulare *PTMA* (*prothymosin alpha*);
- *TERT* (*telomerase reverse transcriptase*);
- Il regolatore del ciclo cellulare *ID2* (*inhibitor of DNA binding 2*);
- *MCM7* (*minichromosome maintenance complex component 7*);
- *IGFIR* (*insulin-like grow factor 1 receptor*).

1.1.2 Amplificazione vs. sovraespressione di MYCN

MYCN è un proto-oncogene; la sua sovraespressione è comunemente, ma non esclusivamente, dipendente da amplificazione genica, ed è stata dimostrata in varie tipologie di tumori. In particolare nel NBL la sovraespressione di *MYCN*, dovuta nel 25-30% dei casi ad amplificazione genica, rappresenta un marcatore di prognosi negativa.

L'importanza di *MYCN* nell'oncogenesi del NBL e potenzialmente di altri tumori - associata alla mancanza di espressione di *MYCN* nelle cellule adulte non maligne - lo rende un importante target per lo sviluppo di nuove terapie farmacologiche, ma anche un interessante marcatore prognostico della forma neoplastica associata.

2. Scopo della tesi

L'espressione del gene *MYCN* è un importante indicatore della severità del NBL.

Poiché il numero di copie di *MYCN* è un indice grezzo della sua espressione quantificarle utilizzando tecniche come la "fluorescent in situ hybridization" (FISH) può servire a formulare una stima del livello di espressione *MYCN* [Shapiro 1993].

Tuttavia, l'espressione aberrante di *MYCN* nel NBL non è sempre associata all'amplificazione genica; pertanto la valutazione diretta del livello di espressione di questo gene sarebbe un miglior indicatore prognostico.

Questa tesi è stata sviluppata nell'ambito di un progetto che si propone di realizzare un sensore biomolecolare sintetico per l'identificazione del livello di espressione di *MYCN*.

Di seguito saranno presentati i dettagli relativi alla progettazione della topologia circuitale e all'analisi *in silico* che sono state condotte per caratterizzare il comportamento dinamico del sistema.

Questo lavoro è stato svolto nel laboratorio di Ingegneria Cellulare e Molecolare "S. Cavalcanti", presso la Sede di Cesena del Dipartimento di Ingegneria dell'Energia elettrica e dell'Informazione "Guglielmo Marconi" (DEI) dell'Ateneo di Bologna.

3. Materiali e Metodi

Per progettare il circuito genetico sintetico che realizza il sensore molecolare di MYCN, è stato utilizzato l'approccio bottom-up tipico della biologia sintetica.

Occorre qui precisare che in questa fase iniziale del progetto, il circuito genetico viene considerato operante in un sistema semplificato, ovvero in cellule batteriche di *E. Coli*.

3.1 Componenti molecolari del circuito (*standard biological parts*)

Elementi di base per la realizzazione del circuito sono parti biologiche identificate da sequenze definite di DNA, caratterizzate sotto il profilo funzionale. Lo standard di riferimento utilizzato per questa fase è quello condiviso dai biologi sintetici che sono attivi in iGEM [http://igem.org/Main_Page], una comunità indipendente e *no-profit* che si dedica all'educazione e allo sviluppo della biologia sintetica, sviluppando una piattaforma collaborativa internazionale. iGEM mantiene un deposito di *Standard Biological Parts*, ovvero una collezione in espansione di componenti genetiche - dette *BioBricks* - che possono essere interfacciate tra di loro con un approccio modulare, per costruire dispositivi molecolari come il sensore oggetto di questo lavoro di tesi. Di seguito una breve descrizione della componentistica considerata per il sensore di MYCN.

3.1.1 Promotori

Un *promotore* è una sequenza di DNA alla quale l'*RNA polimerasi* si lega

per iniziare la trascrizione del gene di competenza. La specifica funzionale di ognuno di essi è la loro forza trascrizionale, ovvero il parametro che definisce quante copie di RNA messaggero verranno sintetizzate nell'unità di tempo. I promotori sempre attivi, quelli ai quali l'RNA polimerasi è libera di legarsi, si dicono *costitutivi*. Ognuno avrà uno specifico rateo cinetico, indicativo della probabilità di trascrizione. I promotori per i quali l'accesso dell'RNA polimerasi è controllato da ulteriori sequenze che si dicono operatori, si definiscono *regolati*. In questo caso lo specifico rateo cinetico è limitato allo stato di attivazione.



Figura 2 Simbologia del promotore costitutivo secondo la schematizzazione di iGEM (<http://partsregistry.org>) (a sinistra) e sua rappresentazione grafica usata in questo elaborato (a destra).

3.1.2 Terminatori

I terminatori sono sequenze di DNA che identificano la regione di un gene dove la polimerasi deve arrestarsi nella trascrizione.

Generalmente si tratta di una sequenza con palindromi che permette il ripiegamento del DNA su se stesso, a formare una struttura a forcina che impedisce fisicamente il proseguimento del percorso della RNA polimerasi lungo il DNA.

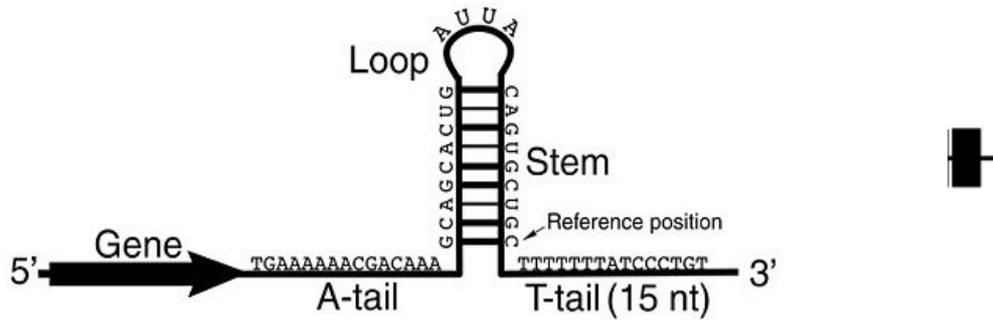


Figura 3 Schematizzazione di un terminatore Rho-indipendente: si può notare come il terminatore impedisce, fisicamente, l'avanzamento della RNA polimerasi che dovrebbe proseguire in direzione 5'-3' nel gene. Sulla destra [<http://parts.igem.org/Terminators>] è riportata la simbologia che verrà nel seguito dell'elaborato.

3.1.3 Siti di legame al ribosoma (RBS)

Gli RBS sono sequenze segnale presenti nel DNA che vengono trascritte in mRNA ma non tradotte. Esse coadiuvano il legame tra l'mRNA e i ribosomi. Sulla base di precedenti analisi sperimentali condotte nel lab. ICM la sequenza RBS considerata nel presente studio *in silico* è quella utilizzata in [Ceroni et al 12] e presenta un'alta affinità con i ribosomi batterici. Nel parts registry è catalogata come BBa_B0034.

3.1.4 Sequenze codificanti

Le sequenze codificanti sono le regioni di DNA che codificano per una determinata proteina. Nel circuito in esame abbiamo unicamente il tratto codificante per il reporter fluorescente verde la Green Fluorescent Protein. Fornendo un readout fluorescente questa proteina è comunemente adottata come reporter dell'attività di circuiti genetici sintetici *in vitro*

3.1.5 Chassis

Come anticipato, il circuito verrà inizialmente implementato in procariote, nello specifico nel batterio E. coli. La veicolazione del circuito genetico nel compartimento cellulare sarà realizzata ricorrendo ad un vettore plasmidico.

I plasmidi sono molecole circolari di DNA extracromosomico a doppio filamento che possiedono un'origine di replicazione (ORI) e sono soggette a duplicazione autonoma.

L'ORI, sequenza responsabile della replicazione del plasmide durante il ciclo cellulare, determina il numero medio di plasmidi presenti in media all'interno della cellula.

Oltre all'ORI, un plasmide è caratterizzato dalla presenza di un gene codificante per la resistenza ad un particolare antibiotico specifico e da un sito multiplo di clonaggio.

Quando sottoposti ad una pressione selettiva, esercitata dalla presenza di antibiotico nel terreno di coltura, solo i trasformanti che ritengono la molecola plasmidica saranno in grado di proliferare.

3.2 Standard assembly

L'adozione di standard fisici comuni per le parti biologiche sopra descritte semplifica e uniforma il clonaggio molecolare di circuiti genetici sintetici. Lo Standard Assembly 10 (Figura 4) prevede l'utilizzo di soli 4 enzimi di restrizione: EcoRI, XbaI, SpeI, PstI. Tali endonucleasi sito specifiche catalizzano la digestione della molecola a doppio filamento di DNA generando estremità coesive, dalle quali origina specificità di appaiamento.

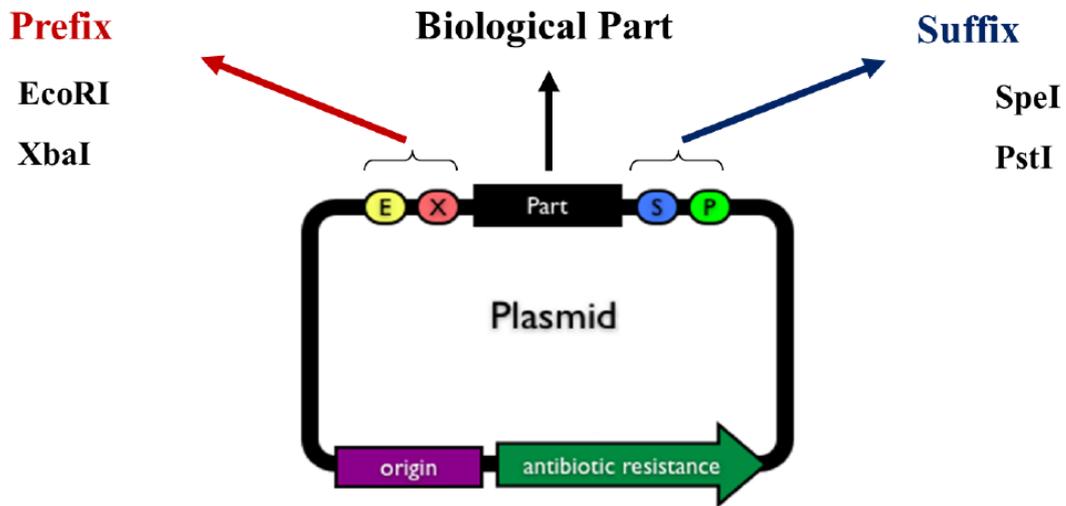


Figura 4: Schema funzionale dello Standard BioBrick Assembly 10 dove sono visibili lo scheletro plasmidico e il sito multiplo di clonaggio flankato da prefisso e suffisso.

Poiché XbaI e SpeI sono isocaudameri (figura 5), ovvero producono estremità coesive complementari e il cui appaiamento genera una scar non più riconoscibile dal set di enzimi in esame, l'assemblaggio di circuiti genetici risulta essere modulare.

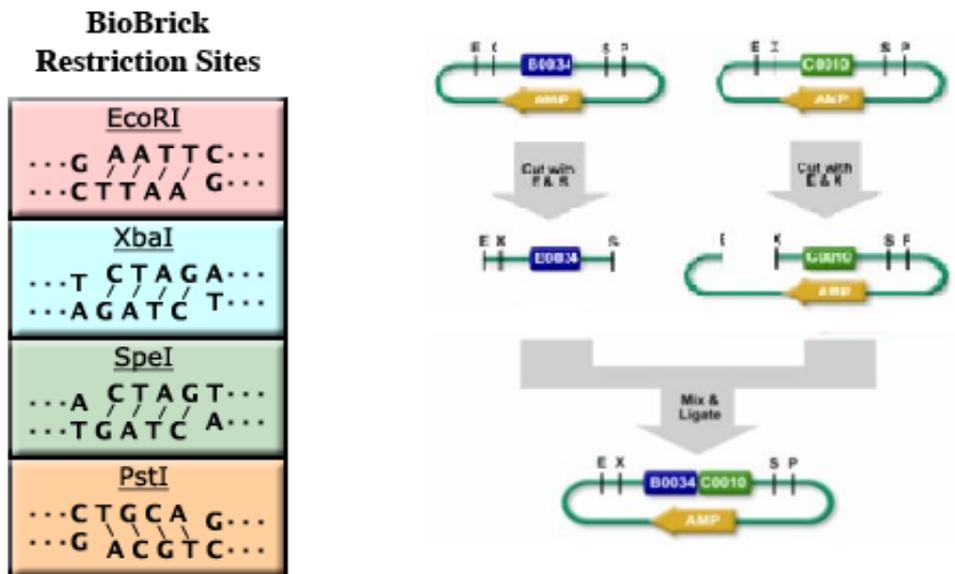


Figura 5: A sinistra troviamo la legenda dei siti di restrizione usati nello standard Assembly 10 e a destra una schematizzazione di uno step del clonaggio molecolare: la digestione E/S ed E/X, per i plasmidi di sinistra e destra rispettivamente, permette l'isolamento e la successiva ligazione di inserto e vettore.

3.3 Topologia e principio di funzionamento del circuito che implementa il sensore molecolare

La topologia del sensore molecolare caratterizzato *in silico* nel presente elaborato, deriva da un setup sperimentale precedentemente sviluppato nel lab. ICM e descritto in (Ceroni 2012). Tale sistema realizza una regolazione post-trascrizionale dell'espressione genica, tramite ibridazione di sue sequenze ribonucleotidiche di regolazione. La prima, denominata CIS e trascritta co-cistronicamente alla sequenza codificante per GFP, si compone di 55 nucleotidi estratti nel contesto dell'esone 2 di *MYCN*, e dell'RBS BBA_B0034. La seconda è complementare al ribonucleotide CIS e verrà indicata con TRANS. A seguito della trascrizione, l'annealing tra i due RNA causa una parziale sovrapposizione (4 bp) all'RBS, inibendo il legame del ribosoma e conseguentemente la traduzione di GFP (Figura 6)

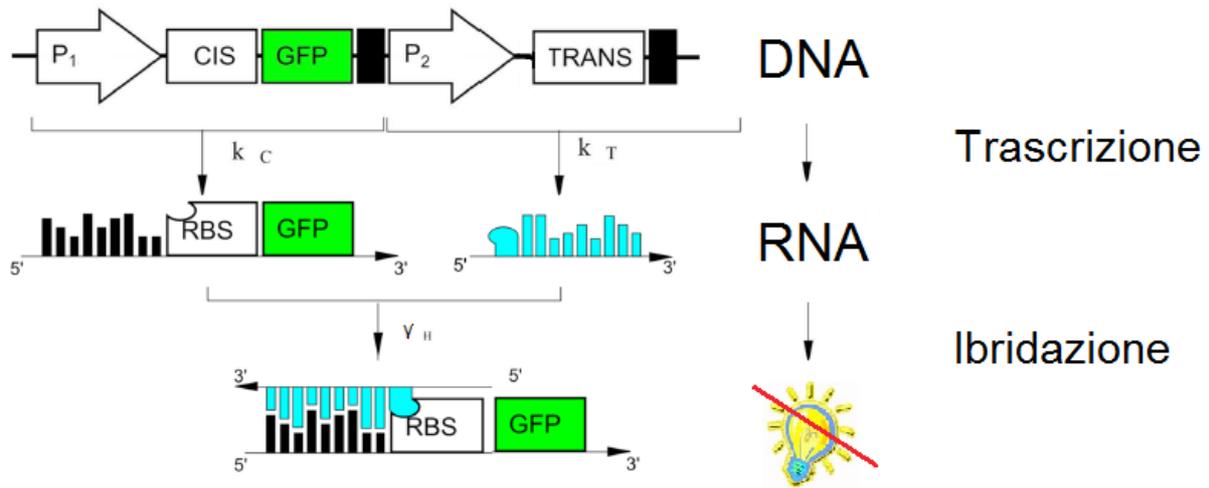


Figura 6: Schema funzionale del sensore biomolecolare. La trascrizione dell'mRNA CIS-GFP e della sequenza ribonucleotidica TRANS è guidata da due promotori costitutivi. L'ultimo step esemplifica il legame tra le due sequenze e l'inibizione della traduzione del reporter fluorescente.

La presenza di trascritto di MYCN, parzialmente coincidente con il ribonucleotide CIS, instaura un meccanismo di competizione per il legame con la sequenza TRANS. Ciò realizza le condizioni per la sintesi di una concentrazione di GFP proporzionale alle copie del trascritto da rilevare (Figura 7).

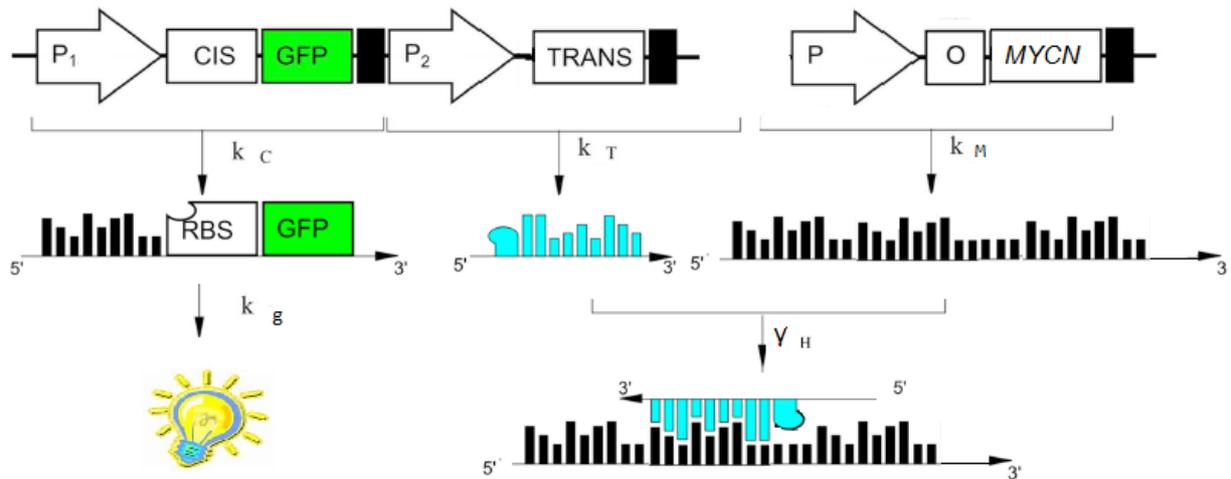


Figura 7: Schema funzionale del sensore molecolare. La competizione tra CIS e MYCN per l'ibridazione con la sequenza TRANS, porta ad un segnale fluorescente proporzionale alla concentrazione di MYCN nel contesto cellulare.

Data l'assenza del gene MYCN nel genoma di *E. coli* il funzionamento del sensore in procariote è subordinato alla sua espressione eterologa attraverso un ulteriore vettore con a bordo il costrutto presentato in Figura 8.

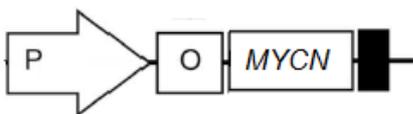


Figura 8: Circuito per l'inserimento dei MYCN all'interno di *E. coli*. A valle del promotore troviamo il sito operatore per la sua regolazione dal momento che prevediamo di produrre diverse quantità di mMYCN.

3.4 Modellazione della dinamica circuitale

Il funzionamento del circuito è stato formalizzato matematicamente

mediante un modello deterministico continuo strutturato in un set di quattro equazioni differenziali. (ODE). Ai fini dell'identificazione dei ratei trascrizionali necessari per un corretto funzionamento del sensore, tali ODE sono state risolte numericamente in MATLAB SIMULINK, a partire da condizioni iniziali nulle, per un tempo sufficiente al raggiungimento dello stato stazionario (SS).

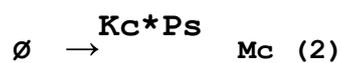
3.5 Reazioni ed equazioni del sistema

Il sistema in esame è stato descritto con 10 reazioni, di cui tre rappresentano la trascrizione dei componenti del sensore molecolare e altre tre descrivono la loro degradazione.

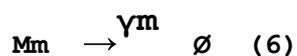
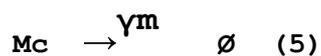
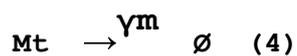
Il sistema si completa con due reazioni di ibridazione, descrittive l'annealing tra mCIS:mTRANS e mTRANS:mMYCN rispettivamente e traduzione del reporter fluorescente.

Il sistema appena descritto è riportato in seguito.

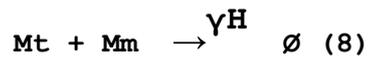
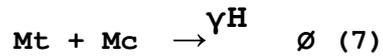
➤ 3 reazioni di trascrizione:



➤ 3 reazioni di degradazione dell'mRNA:



➤ 2 reazioni di ibridazione:



➤ 1 reazione di traduzione:



➤ 1 reazione di degradazione di proteina:



Da queste reazioni è stato ricavato il corrispondente sistema di ODE, che è stato simulato per predire sensitività e range dinamico del sensore molecolare.

➤ $dM_t/dt = K_t \cdot P_s - \gamma_m \cdot M_t - \gamma^H \cdot M_t \cdot M_c - \gamma^H \cdot M_t \cdot M_m \quad (1)$

➤ $dM_c/dt = K_c \cdot P_s - \gamma_m \cdot M_c - \gamma^H \cdot M_t \cdot M_c \quad (2)$

➤ $dM_m/dt = K_m \cdot M_m - \gamma_m \cdot M_m - \gamma^H \cdot M_t \cdot M_m \quad (3)$

➤ $dG/dt = K_g \cdot M_c - \gamma_g \cdot G \quad (4)$

Come solutore per il modello implementato in SIMULINK si è scelto ODE23s a seguito dello studio della stiffness del problema che è stata quantificata valutando l'indice di stiffness r a partire dal determinante dello Jacobiano in condizioni iniziali:

$$J_0 = \begin{pmatrix} -\gamma m & 0 & 0 & 0 \\ 0 & -\gamma m & 0 & 0 \\ 0 & 0 & -\gamma m & 0 \\ 0 & kg & 0 & -\gamma g \end{pmatrix}$$

L'indice di stiffness è definito come:

$$r = \max(|\Lambda|) / \min(|\Lambda|);$$

dove Λ è la matrice degli autovalori dello Jacobiano all'istante iniziale (J_0).

Il valore calcolato per r (9.3), eccede la soglia unitaria oltre la quale un problema è definito stiff.

3.6 Parametri del modello

I parametri e le variabili di stato del modello sono riportate nella tabella seguente:

Simbolo	Descrizione	Valore / range in cui è stata fatta l'analisi	Unità di misura	riferimento
Mt	mTRANS	/	Molecole per cellula	
Mc	MCISGFP	/	Molecole per cellula	
Mm	mMYCN	/	Molecole per cellula	
G	GFP	/	Molecole per cellula	
Kt	Rateo trascrizionale	0,01-1	min ⁻¹	http://

	di TRANS			parts.i-gem.org/Promoters/Catalog/Anderson
Kc	Rateo trascrizionale di CIS	0,01-1	min ⁻¹	
Km	Rateo trascrizionale di MYCN	0-1	min ⁻¹	
γH	Rateo di ibridazione degli mRNA	0-10	(molecole*min) ⁻¹	
Kg	Rateo di traduzione di GFP	11,5	min ⁻¹	[Ceroni 12]
γm	Rateo di degradazione degli mRNA	0,2	min ⁻¹	[Bakshi et al 2012]
γg	Rateo di degradazione della GFP	0,0215	min ⁻¹	[Ceroni 12]
Ps	Copy number del plasmide su cui è clonato il circuito genetico	10	Plasmidi per cellula	
Pm	Copy number del plasmide su cui è montato il circuito genetico per la produzione mMYCN	5	Plasmidi per cellula	

Lo sfondo azzurro identifica le variabili di stato, quello arancio i parametri fissati o noti mentre in rosa sono contrassegnati i parametri che verranno variati durante l'analisi.

Tenendo presente che l'emivita di una molecola corrisponde al tempo necessario affinché la sua concentrazione si dimezzi, il rateo di degradazione dell'mRNA può essere calcolato con la seguente formula:

$$\gamma_m = \ln(2)/\tau = 0,20 \text{ min}^{-1}$$

dove per l'emivita (τ) è è assunto un valore intermedio al range desunto da letteratura scientifica [Bakshi et al 2012].

Non potendo identificare a priori il valore del rateo di ibridazione , l'effetto di questo parametro sul funzionamento del circuito è stato valutato nel range [0:10].

Al fine di svolgere simulazioni in un contesto in linea con la futura realizzazione fisica del circuito sintetico, il numero medio di plasmidi veicolanti il sensore il sensore e il generatore di MYCN sono stati fissati a 10 e 5 rispettivamente. Tali valori sono compatibili con un vettore plasmidico a basso numero di copie che risulta ben tollerato dalle cellule batteriche dal punto di vista metabolico.

3.7 identificazione in silico dei ratei trascrizionali

Per quanto riguarda i promotori, si è deciso di limitare la scelta a quelli contenuti nella libreria di Anderson [<http://parts.igem.org/Promoters/Catalog/Anderson>] articolata in 18 elementi la cui forza trascrizionale è stata caratterizzata ed è riportata, nella tabella sottostante, in rapporto al rateo trascrizionale del promotore wild-type (BBa_J23100).

Indentificativo	Sequenza genetica	Forza relativa del promotore
BBa_J23100	ttgacggctagctcagtcctaggtacagtgctagc	1
BBa_J23101	tttacagctagctcagtcctaggtattatgctagc	0.70
BBa_J23102	ttgacagctagctcagtcctaggtactgtgctagc	0.86
BBa_J23103	ctgatagctagctcagtcctagggattatgctagc	0.01
BBa_J23104	ttgacagctagctcagtcctaggtattgtgctagc	0.72
BBa_J23105	tttacggctagctcagtcctaggtactatgctagc	0.24
BBa_J23106	tttacggctagctcagtcctaggtatagtgctagc	0.47
BBa_J23107	tttacggctagctcagccctaggtattatgctagc	0.36
BBa_J23108	ctgacagctagctcagtcctaggtataatgctagc	0.51
BBa_J23109	tttacagctagctcagtcctagggactgtgctagc	0.04
BBa_J23110	tttacggctagctcagtcctaggtacaatgctagc	0.33
BBa_J23111	ttgacggctagctcagtcctaggtatagtgctagc	0.58

<u>BBa_J23112</u>	ctgatagctagctcagtcctagggattatgctagc	0.00
<u>BBa_J23113</u>	ctgatggctagctcagtcctagggattatgctagc	0.01
<u>BBa_J23114</u>	tttatggctagctcagtcctaggtacaatgctagc	0.10
<u>BBa_J23115</u>	tttatagctagctcagcccttgggtacaatgctagc	0.15
<u>BBa_J23116</u>	ttgacagctagctcagtcctagggactatgctagc	0.16
<u>BBa_J23117</u>	ttgacagctagctcagtcctagggattgtgctagc	0.06
<u>BBa_J23118</u>	ttgacggctagctcagtcctaggtattgtgctagc	0.56

Al fine di svolgere simulazioni in un contesto in linea con la futura realizzazione fisica del circuito sintetico, il numero medio di plasmidi veicolanti il sensore il sensore e il generatore di MYCN sono stati fissati a 10 e 5 rispettivamente. Tali valori sono compatibili con un vettore plasmidico a basso numero di copie che risulta ben tollerato dalle cellule batteriche dal punto di vista metabolico.

La valutazione dei componenti biologici utili alla implementazione fisica del sensore biomolecolare, richiede la scelta di un valore soglia di mMYCN, rilevante dal punto di vista prognostico.

L'espressione di mMYCN esclusivamente in conseguenza di un fenotipo patologico, ha portato, in accordo con personale clinico, 10 molecole di trascritto come indicatore prognostico.

Il burden metabolico associato alla sintesi di GFP rende una concentrazione superiore a 3000 proteine/cellula irrealistica [Roberts et al 2011], Tale considerazione, nel seguito definita vincolo di saturazione, ha portato all'esclusione delle coppie di promotori per i quali il valore stazionario di reporter fluorescente superava questo limite.

Una volta valutato che il livello di GFP fosse fisiologico, si è proceduto all'analisi del range dinamico ottenibile nelle diverse configurazioni. Si è pertanto deciso di selezionare solo le coppie di promotori responsabili di una marcata differenza nella sintesi di GFP tra le condizioni:

- i) mMYCN= 10 molecole;
- ii) mMYCN= 0 molecole.

In particolare si è imposto che la concentrazione di GFP sintetizzata in assenza di mMYCN, , non eccedesse il 10% del valore ottenuto in

presenza di un numero di copie di trascritto pari a 10.

Al fine di assicurare la capacità di discriminare sperimentalmente i segnali fluorescenti associati a un numero variabile di copie del trascritto, si è deciso di escludere le coppie di promotori che non eccedessero le 1000 unità di GFP in corrispondenza del valore soglia di mMYCN.

In Figura 9 una mappa di calore è adottata per rappresentare l'espressione differenziale del reporter fluorescente nelle configurazioni menzionate in precedenza.

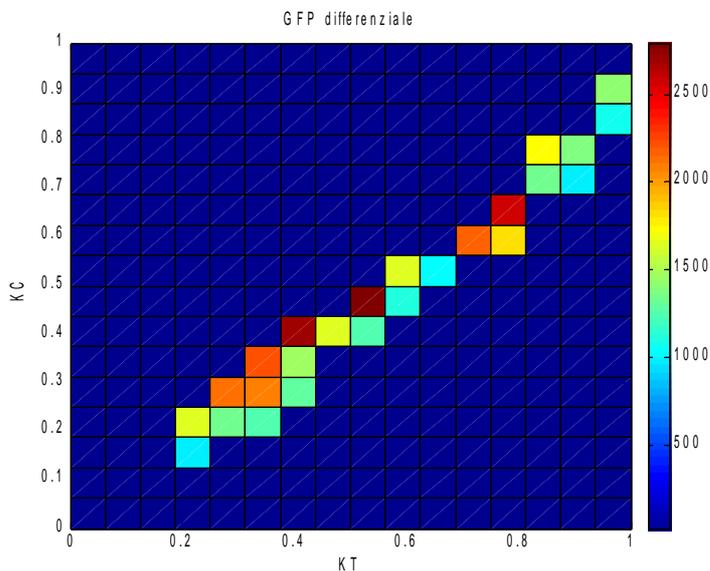


Figura 9: Rappresentazione dell'espressione differenziale di GFP allo stato stazionario. Le coppie di promotori responsabili di più elevate variazioni del segnale fluorescente supportano una più semplice detezione del valore prognostico adottato come riferimento.

Alle migliori coppie di promotori è stato successivamente assegnato uno score rappresentativo della loro capacità di implementare un corretto funzionamento del sensore molecolare al variare del rateo di ibridazione.

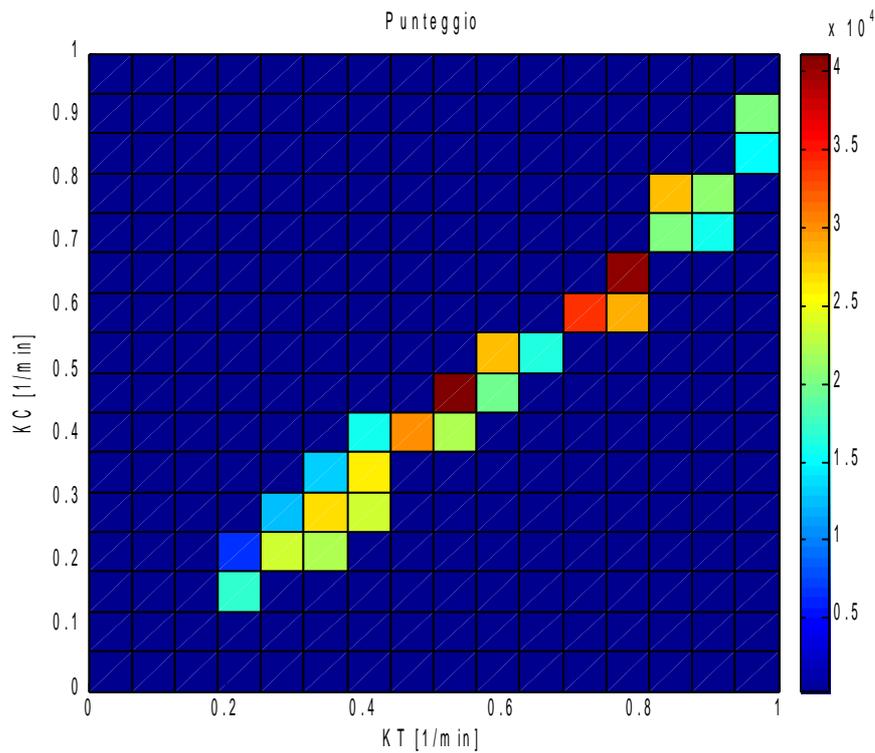


Figura 10: Punteggio delle coppie di promotoriali. Pur mantenendo coerenza con i risultati mostrati in Figura 11, la selezione basata sullo score restringe lo spazio delle coppie di promotori di interesse.

Questa selezione ha permesso di identificare l'1% di coppie maggiormente significative, che, dopo essere state sottoposte a una più approfondita analisi *in silico*, verranno proposte per la realizzazione fisica del sensore.

Le sezioni successive dettagliano i risultati dell'analisi numerica svolta in questo lavoro di tesi, presentandone alcune conclusioni.

4. Risultati e discussione

4.1 Identificazione del tratto (~ 50 nt) di sequenza MYCN adeguata per realizzare i blocchi CIS e TRANS del sensore molecolare

le specifiche sequenze CIS e TRANS del sensore molecolare sono state definite a seguito di analisi dell'intera sequenza del gene, disponibile su http://www.ncbi.nlm.nih.gov/nuccore/NM_005378.4 basato su una prima condizione da soddisfare per la fattibilità del dispositivo è stata l'utilizzo di sequenze di DNA effettivamente trascritte. Per questa ragione è stato necessario escludere dalla selezione il primo dei 3 esoni componenti il gene perché esso ha siti promotori alternativi con corrispondenti sedi alternative di inizio della trascrizione dell'esone 1. i successivi esoni sono invece trascritti integralmente e sono quindi stati oggetto dell'analisi.

Per rendere massima la specificità del blocco TRANS del sensore, non sono state prese in considerazione le regioni di MYCN che risultassero parzialmente omologhi al genoma umano e di E. Coli. Questo confronto è stato possibile grazie all'algoritmo BLAST® disponibile a <http://blast.ncbi.nlm.nih.gov>, che è in grado di individuare regioni di similarità di sequenze nucleotidiche (o di amminoacidi) calcolando un indice di significatività statistica. Sono ugualmente state scartate le sequenze che presentavano significative probabilità di ibridazione intramolecolari. Infine si è scelta una sequenza che avesse il minor numero di ripetizioni di CG. Le regioni ricche di guanina e citosina producono complessi ripiegamenti intra- e intercatena (forcine e anse) conseguenti alla formazione di multipli legami idrogeno fra nucleotidi adiacenti (Musso et al 2006). Per garantire una sufficiente energia di ibridazione CIS/TRANS sono state

considerate utilizzabili sequenze di lunghezza pari a 55 nucleotidi.

Sono state dunque individuate 2 regioni di MYCN, riportate di seguito nella direzione 5'-3', che rispettassero i requisiti prima elencati:

GGAAGGAAGCACCCCCGGTATTAAAACGAACGGGGCGGAAAGAAGCCCTCAGTCGCCGGCCGGGAGGCGA
GCCGATGCCGAGCTGCTCCACGTCCACCATGCCGGGCATGATCTGCAAGAACCCAGACCTTCGAGTTTGAC
TCGCTACAGCCCTGCTTCTACCCGGACGAAGATGACTTCTACTTTCGGCGGCCCCGACTCGACCCCCCGG
GG

e

CTACGTGGAGAGTGAGGATGCACCCCCACAGAAGAAGATAAAGAGCGAGGCGTCCCCACGTCCGCTCAAG
AGTGTCATCCCCCAAAGGCTAAGAGCTTGAGCCCCCGAAACTCTGACTCGGAGGACAGTGAGCGTTCGCA
GA

la regione sottolineata nella prima regione rappresenta quella selezionata per sintetizzare CIS.

5' TCGAGTTTGACTCGCTACAGCCCTGCTTCTACCCGGACGAAGATGACTTCTACTT 3'

Dunque la sequenza TRANS sarà la sua complementare antiparallela

5' AAGTAGAAGTCATCTTCGTCCGGGTAGAAGCAGGGCTGTAGCGAGTCAAACCTCGA 3'

4.2 identificazione dei parametri

Secondo il procedimento spiegato nel capitolo precedente, si è simulato in SIMULINK il modello deterministico continuo, strutturato in un sistema di quattro equazioni differenziali, ai fini della valutazione del numero di molecole di Mmycn e di GFP presenti nel compartimento cellulare in stato stazionario (SS).

Le simulazioni sono state svolte assegnando in maniera combinatoria gli elementi della libreria di Anderson [<http://parts.igem.org/Promoters/Catalog/Anderson>] ai promotori del circuito genetico responsabili della trascrizione di CIS e TRANS. Mentre per il generatore di mMYCN, sono state considerati 25 possibili ratei di trascrizione equi spaziatati e compresi fra 0 e 1. Per quanto riguarda le costanti di ibridazione sono stati considerati 20 valori equidistanti tra 0 e 10 .

Di mMYCN in realtà è rilevante conoscere solo i suoi livelli di produzione al variare dei promotori, dal momento che è questo che ci interessa rilevare la sua presenza totale all'interno della cellula. Ne consegue che di mMYCN in SS è considerato il vettore a ibridazione nulla.

Identificati i valori di mMYCN, si è cominciato con il verificare quali coppie rispettassero i requisiti sulla produzione di GFP a mMYCN pari a 10 copie, ovvero che il circuito genetico sintetizzasse un numero di molecole di GFP compreso tra 1000 e 3000 e che invece in assenza di mMYCN GFP non fosse superiore al 10% del valore individuato al valore soglia di mMYCN.

In questo modo si è ridotto lo spazio delle coppie promotoriali di interesse.

Fra le coppie che soddisfano i requisiti però, è stato necessario fare un'ulteriore selezione, sfruttando due fattori come pesi che sarebbero andati a creare un punteggio per la coppia di promotori.

Della coppia infatti sono interessanti due proprietà, sulle quale si è concentrata l'analisi:

- Differenza di GFP fra condizione in assenza di mMYCN e presenza di 10 mMYCN;
- Stabilità del comportamento al variare del rateo di ibridazione.

Per creare il punteggio, si è deciso di sfruttare la GFP differenziale fra i due stati di interesse sommando questo valore ogni volta che la coppia rispettava i requisiti imposti in precedenza in funzione del

rateo di ibridazione.

In questo modo, è stata creata la matrice dei punteggi che viene riportata di seguito riproposta come mappa di calore.

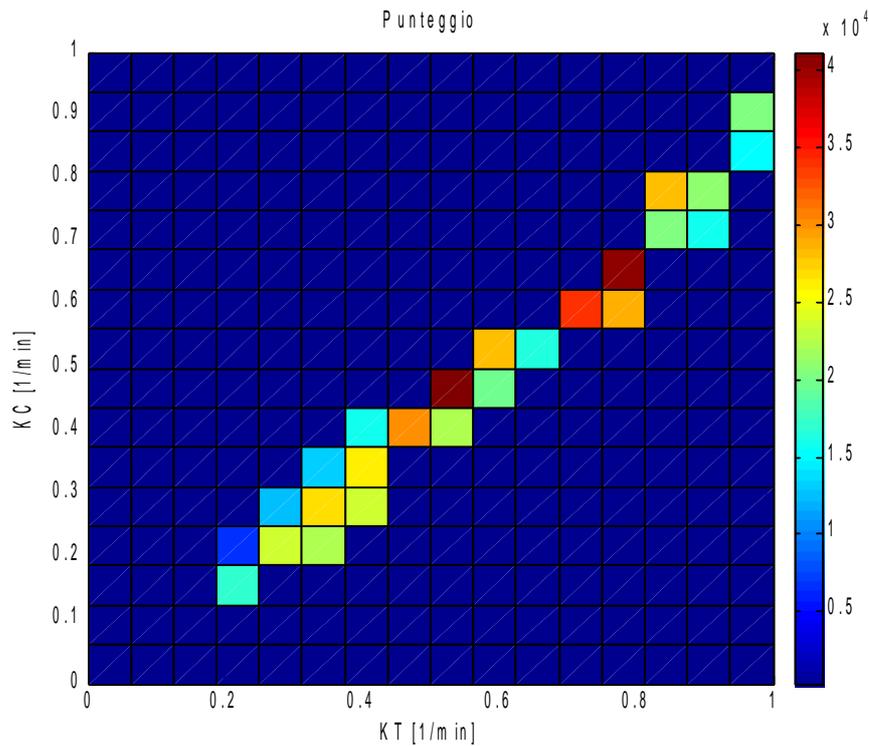


Figura 12: Mappa di calore rappresentante il punteggio delle varie coppie di promotori, ottenuto considerando sia la GFP differenziale, sia la permanenza della validità della coppia all'interno del range considerato della costante di ibridazione. Le coppie rosse scuro risultano le coppie col punteggio maggiore e che saranno oggetto della prossima analisi.

Innanzitutto si nota che il range delle coppie che rispettano i requisiti è molto ristretto rispetto alla totalità delle coppie e si concentra su forze promotoriali simili fra i due promotori.

Per proseguire con l'analisi, si è deciso di concentrarsi sulle sole coppie di promotori che avessero un punteggio superiore a 3×10^4 , in modo da avere una visione più ampia delle coppie e non solo delle due coppie con punteggio maggiore che si vedono in rosso scuro nella figura 11.

Di seguito sono emersi i seguenti andamenti di GFP in funzione di mMYN

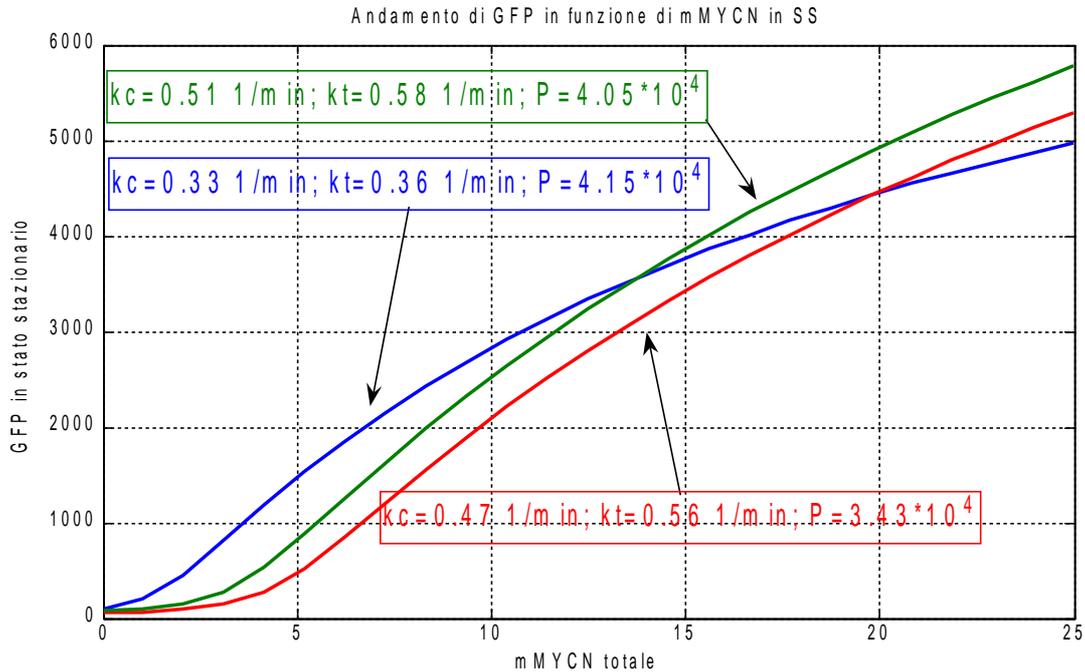


Figura 13: Andamento della GFP in SS al variare della quantità di mMYCN per le coppie di promotori che hanno il punteggio maggiore, confrontate per lo stesso valore della costante di ibridazione. Possiamo notare che il in corrispondenza del valore 10 mMYCN, le curve con punteggio maggiore sono più alte rispetto a quelle con punteggio minore ottenendo dunque un maggiore range dinamico.

Di queste tre coppie di promotori, abbiamo valutato l'andamento della GFP in funzione di mMYCN al variare del rateo di ibridazione.

Per tutte le coppie notiamo che all'aumentare del rateo di ibridazione la concentrazione di GFP a mMYCN nullo diminuisce, portando ad un aumento del range dinamico.

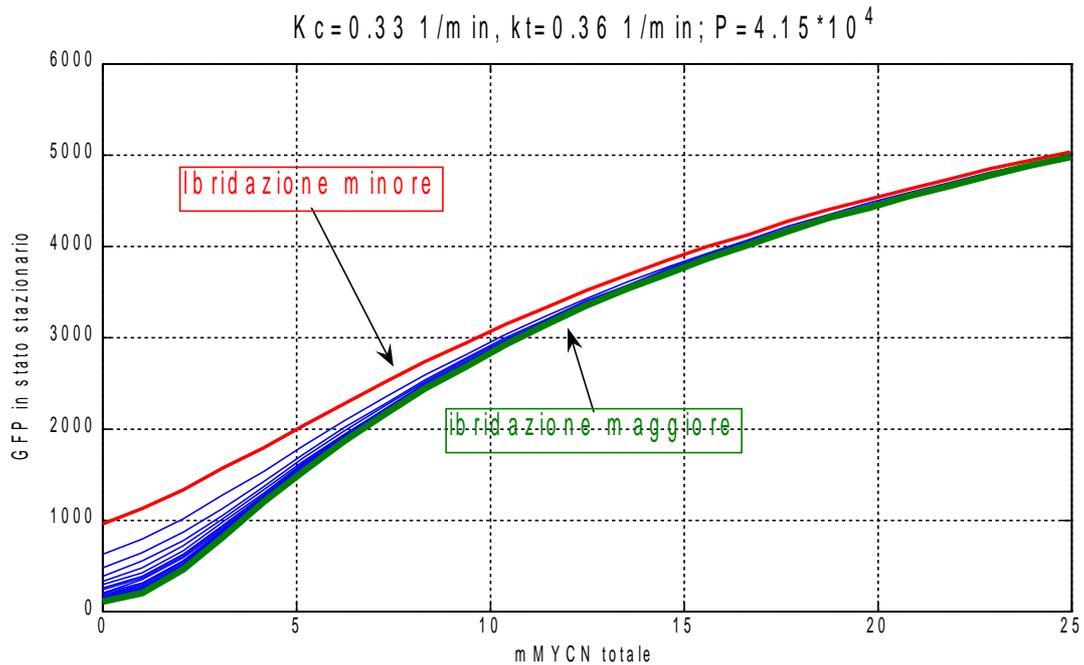


Figura 14: Nel grafico sono riportati tutti gli andamenti della GFP in SS al variare dei livelli di mMYCN per la coppia $K_c = 0,33 \text{ 1/min}$ e $K_t = 0,36 \text{ 1/min}$ al variare della costante di ibridazione. Possiamo notare come aumenti il range dinamico all'aumentare dell'ibridazione in conseguenza allo spostamento della curva verso il basso in maniera più significativa per valori bassi di mMYCN rispetto che nei valori più alti, dove invece i livelli raggiunti sono per lo più invariati.

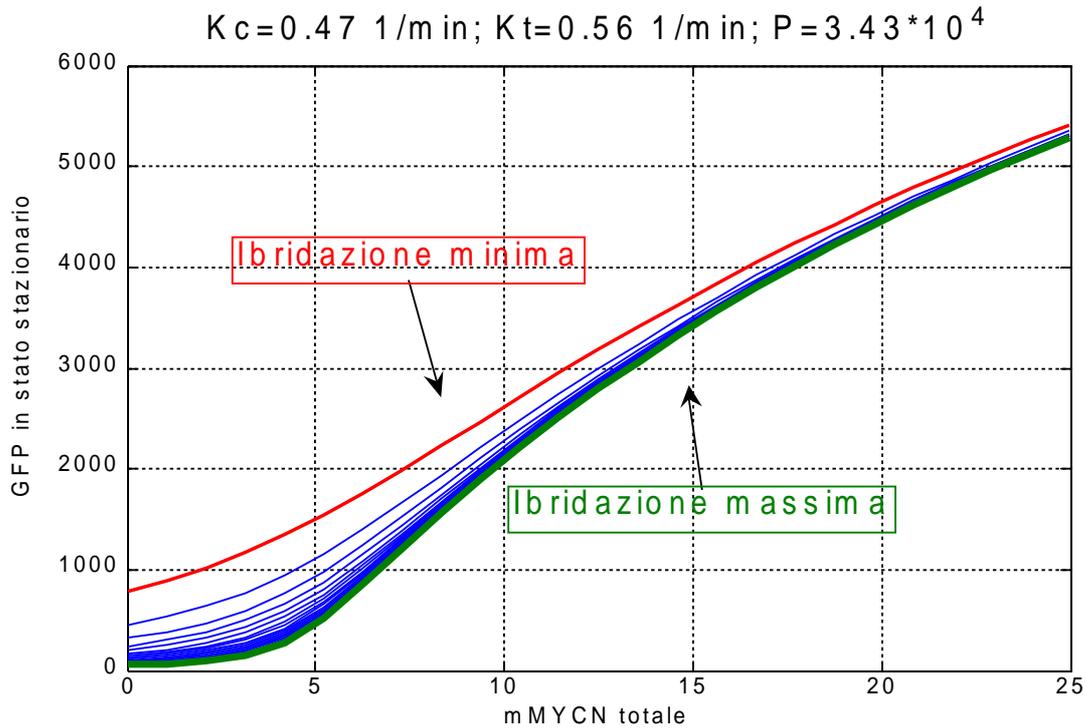


Figura 15: Andamento della GFP in SS al variare di mMYCN per la coppia $K_c=0,47 \text{ 1/min}$ e $K_t=0,56 \text{ 1/min}$ al variare del rateo di ibridazione. Come nel caso precedente notiamo lo stesso abbassamento della curva all'aumentare della ibridazione.

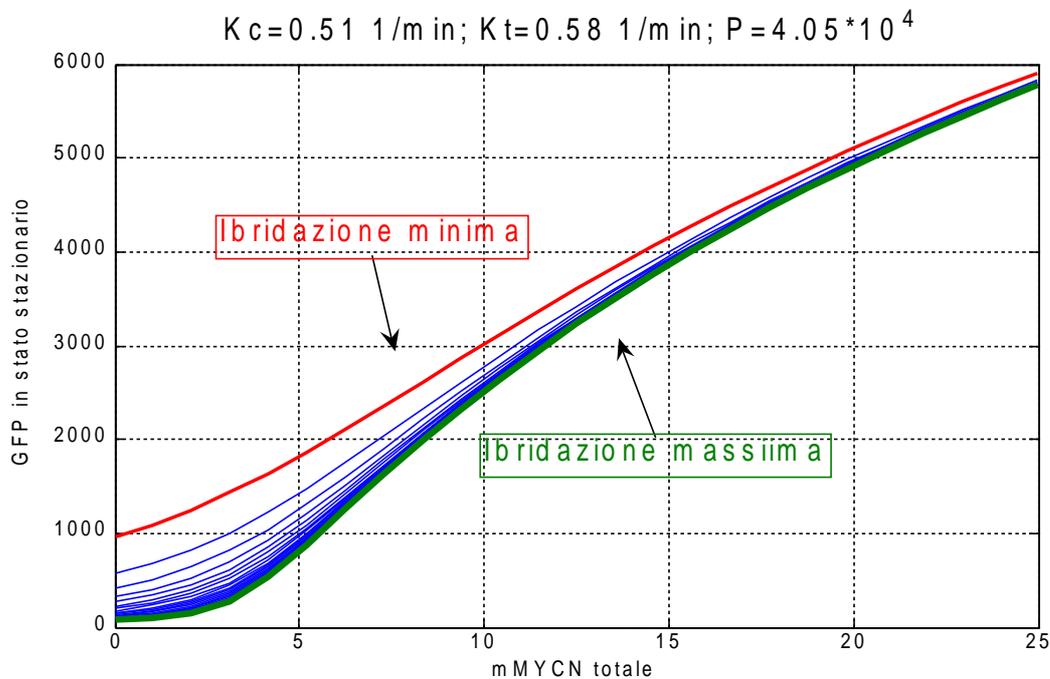


Figura 16: Andamento della GFP in SS al variare di mMYCN per la coppia $K_c=0,51 \text{ 1/min}$ e $K_t=0,58 \text{ 1/min}$ al variare del rateo di ibridazione. Come nel caso precedente notiamo lo stesso abbassamento della curva all'aumentare della ibridazione.

5. Conclusioni

Le migliori sequenze desumibili da quelle del gene MYCN utilizzabili per sintetizzare le sequenze nucleotidiche CIS e TRAN sono risultate:

CIS: 5' TCGAGTTTGACTCGCTACAGCCCTGCTTCTACCCGGACGAAGATGACTTCTACTT 3'

TRANS: 5' AAGTAGAAGTCATCTTCGTCCGGGTAGAAGCAGGGCTGTAGCGAGTCAAACTCGA 3'

Le analisi effettuate in silico tramite MATLAB SIMULINK hanno indicato tre assortimenti di promotori (su oltre 250 possibili utilizzando in maniera combinatoria gli elementi della libreria Anderson) che raggiungono il miglior range dinamico per discriminare in SS il numero di copie (0 - 10) di trascritti di MYCN quando 10 copie del messaggero determinano una produzione di 3000 copie di GFP, inteso come limite massimo fisicamente raggiungibile in E. coli.

Tali coppie promotoriale, ca clonare rispettivamente a monte dei CIS e TRANS nella sonda molecolare, montata in un plasmide con numero di copie pari a 10, risultano le seguenti:

KCIS1 = 0,33 min⁻¹; KTRANS1 = 0,36 min⁻¹;

KCIS2 = 0,47 min⁻¹; KTRANS2 = 0,51 min⁻¹;

KCIS3 = 0,51 min⁻¹; KTRANS3 = 0,58 min⁻¹;

6. Ringraziamenti

Ringrazio il mio relatore Prof. Emanuele D. Giordano per la responsabilità assunta e le mie correlatrici, Ing. Lucia Bandiera e Ing. Marilisa Cortesi e per avermi guidato nello sviluppo della tesi. Senza il loro aiuto non avrei concluso il mio lavoro. In particolare, sono in debito con entrambe per la pazienza con la quale hanno ascoltato i miei dubbi e le mie incertezze.

Un ringraziamento al Dott. Simone Furini, che mi ha fornito preziosi consigli sui procedimenti per l'identificazione di parametri rilevanti per la simulazione numerica.

7. Bibliografia:

- "Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells" S. Bakshi, A. Siryaporn, M. Goulian, JC. Weisshaar, 2012.
- "A Synthetic Post-transcriptional Controller To Explore the Modular Design of Gene Circuits" F. Ceroni, S. Furini, A. Stefan, A. Hochkoeppler, E. Giordano, 2012.
- "Identification and Pharmacological Inactivation of the MYCN Gene Network as a Therapeutic Strategy for Neuroblastic Tumor Cells", Olesya Chayka, Cosimo Walter D'Acunto, Odette Middleton, Maryam Arab and Arturo Sala, THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 290, NO. 4, pp. 2198-2212, January 23, 2015.
- Musso et al. Betaine, dimethyl sulfoxide, and 7-deaza-dGTP, a powerful mixture for amplification of GC-rich DNA sequences. *J Mol Diagn.* (2006) 8:544-550.
- Nesbit et al. "MYC oncogenes and human neoplastic disease" *Oncogene* (1999) 18:3004-3016.
- "MYC oncogenes and human neoplastic disease", Chadd E. Nesbit, Jean M. Tersak and Edward V. Prochownik, *Oncogene* (1999) 18, 3004 - 3016.
- "Inibizione selettiva del gene *MYCN* mediante PNA (acidi peptidico nucleici) anti-gene nel raiomiosarcoma umano", S. Purgato, 2008
- "Noise contributions in an inducible genetic switch: a whole-cell simulation study" E. Roberts, A. Magis, J. O. Ortis, W. Baumeister, Z. Luthey-Shultzen. 2011.
- "Detection of N-myc Gene Amplification by Fluorescence in Situ Hybridization. Diagnostic Utility for Neuroblastoma", David N. Shapiro, Marcus B. Valentine, Susan T. Rowe, Anne E. Sinclair,

Jack E. Sublett, W. Mark Roberts and A. Thomas Look, 1993.

- Shapiro et al. "Detection of N-myc gene amplification by fluorescence in situ hybridization. Diagnostic utility for neuroblastoma" Am J Pathol (1993) 142:1339-1346.
- "NCYM promotes calpain-mediated Myc-nick production in humanMYCN-amplified neuroblastoma cells", Wataru Shoji, Yusuke Suenaga, Yoshiki Kaneko, S.M. Rafiqul Islam, Jennifer Alagu, Sana Yokoi, Masaki Nio, Akira Nakagawara, 2015.
- Stanton et al. "Alternative processing of RNA transcribed from NMYC" Mol Cell Biol (1987) 7:4266-4272.

8. Sitografia:

- <http://ghr.nlm.nih.gov/gene/MYCN>
- <http://parts.igem.org>
- <http://parts.igem.org/Promoters/Catalog/Anderson>
- http://www.ncbi.nlm.nih.gov/nucleotide/NM_005378.4
- <http://parts.igem.org/Terminators>
- <http://www.dei.unibo.it/it/ricerca/strutture/laboratori/icm-laboratorio-di-ingegneria-cellulare-e-molecolare-s.-cavalcanti>
- <http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form>
- <http://blast.ncbi.nlm.nih.gov>