

The seal of the University of Padua is a large, circular emblem in red. It features a central shield divided into two panels. The left panel shows a seated female figure, likely the personification of Justice, holding a scale and a sword. The right panel shows a standing male figure, likely the personification of Reason, holding a book and a staff. The shield is surrounded by a decorative border of small stars. The outer ring of the seal contains the Latin text 'UNIVERSITAS STUDII PADUENSIS' in a stylized font, with a diamond symbol separating the words. At the bottom of the seal, the year 'MCCXII' is inscribed.

TURNOVER DELLE PROTEINE

Laureando
Stefan Octavian Popa

Relatore
Ch.ma Prof.ssa Gianna Maria Toffolo

Indice

1	Introduzione	1
2	Reti di trascrizione	3
2.1	Introduzione	3
2.2	Interazione tra cellule e l'ambiente	3
2.3	Reti di trascrizione	5
2.4	Gli archi nelle reti di trascrizione	10
2.4.1	Funzione di Hill	11
2.4.2	Funzioni logiche	13
2.4.3	Funzioni multidimensionali che regolano i geni con vari input	14
2.5	La regolazione di un gene	14
2.5.1	Tempo di risposta di una proteina stabile	18
3	Autoregolazione	21
3.1	Introduzione	21
3.2	Reti reali e reti random	21
3.3	Autoregolazione: motivi	23
3.4	Vantaggi dell'autoregolazione	24
3.4.1	L'autoregolazione negativa porta robustezza ai flussi di produzione	27
3.4.2	L'autoregolazione positiva	29
4	Regolazione feed forward	31
4.1	Introduzione	31
4.2	Motivi con 3 nodi	31
4.3	Feed forward	32

IV

INDICE

5 Conclusioni

37

Bibliografia

39

Elenco delle figure

2.1	Mappa dei segnali, dei fattori di trascrizione e dei geni che essi regolano	4
2.2	Trascrizione di un gene	5
2.3	Fattore di trascrizione: attivatore	6
2.4	Fattore di trascrizione: repressore	7
2.5	Rete di trascrizione	8
2.6	Funzione di Hill per un fattore di trascrizione attivatore	11
2.7	Funzione di Hill per un fattore di trascrizione repressore	12
2.8	Andamento esponenziale decrescente della concentrazione di una proteina in risposta all'assenza di un segnale di ingresso .	17
2.9	Andamento esponenziale della concentrazione di una proteina in risposta ad una salita a gradino	18
3.1	Confronto tra una rete reale e una rete random	22
3.2	Rete di trascrizione	23
3.3	Confronto tra gene senza autoregolazione (a) e gene con autoregolazione negativa (b)	25
3.4	La dinamica della produzione di un gene autoregolato negativamente	26
3.5	Misurazione sperimentale del motivo di regolazione in una rete reale	28
3.6	Andamento della concentrazione di una proteina: a) autoregolazione negativa, b) assenza di autoregolazione, c) autoregolazione positiva	29
4.1	Possibili combinazioni dei moduli di regolazione con 3 nodi . . .	32
4.2	a) Regolazione feed forward, b) regolazione feedback	32
4.3	Motivo di regolazione feed forward	33

4.4	Andamento temporale delle concentrazioni di Y e Z	34
4.5	Risposta a stimoli del motivo di regolazione feed forward . . .	35

Capitolo 1

Introduzione

Nell'ultimo decennio nell'ambito biomedico è cresciuto l'interesse verso la proteomica¹, per i seguenti motivi:

- 1) Lo studio del proteoma umano ha attirato su di sé una notevole attenzione dopo aver evidenziato che il genoma è costituito da molto meno geni rispetto alle proteine. Di fatto nell'uomo sono stati stimati circa 35 mila geni e tra 250 e 500 mila proteine, di cui solamente una parte è stata identificata e valutata. Inoltre non è sufficiente conoscere le informazioni genomiche per spiegare gli eventi biologici e per individuare nuovi bersagli e terapie farmacologiche.
- 2) I profili dell' mRNA non si correlano sempre con il livello di espressione proteica. Infatti l'analisi del trascrittoma² ci dà un'indicazione di quali siano i geni attivi in una determinata cellula, ma nulla riguardo le proteine effettivamente espresse. La mancanza di equivalenza tra trascrittoma e proteoma è dovuta al fatto che non tutti gli mRNA vengono tradotti in proteine e che il contenuto proteico della cellula è determinato sia dalla velocità di sintesi delle nuove cellule che la velocità di degradazione delle proteine già esistenti.
- 3) In seguito ai progressi tecnologici e alle nuove tecniche di spettrometria di massa è stato possibile spostare l'analisi anche al livello del proteoma, dopo aver studiato il singolo gene e analizzato l'espressione dell'intero genoma.

Gli obiettivi primari della proteomica sono la determinazione delle sequenze amminoacidiche che determinano le proteine, la definizione della loro

¹Scienza che studia l'insieme delle proteine espresse dal genoma di un organismo

²Insieme di tutti i trascritti (RNA messenger) di un dato organismo o tipo cellulare.

struttura tridimensionale e quindi l'individuazione della loro funzione biologica e l'indicazione del modo in cui interagiscono fra di loro e con altre molecole.[2]

Recentemente si sta sviluppando la *proteomica clinica* che si pone l'obiettivo di sviluppare tecnologie per la cura e la diagnosi di malattie attraverso l'analisi e il confronto del proteoma di cellule sane e cellule malate. Ciò permette di individuare proteine nell'organismo che possano essere utilizzate come marker precoci della malattia oppure che possano predire la risposta alla terapia di un dato trattamento.

Questi studi non sono sufficienti, in quanto i livelli di espressione proteica non sempre riflettono tutte le alterazioni indotte nel sistema perché la concentrazione di una proteina misurata nel plasma è il risultato della sua velocità di sintesi e la sua velocità di degradazione. In seguito ad una perturbazione, le velocità dei processi di sintesi e degradazione possono subire una variazione. Se rimangono prossimi all'equilibrio, l'espressione proteica rimane quasi inalterata; se invece ci sono variazioni notevoli nella velocità di sintesi di una proteina³, queste non sempre vengono riflesse sulla variazione del livello di espressione proteica. Questo può essere uno dei motivi per cui c'è una scarsa correlazione tra i valori di espressione di geni e proteine corrispondenti, come riportato in letteratura per il lievito *Saccharomyces cerevisiae* (Gygi et al, 1999, Griffi net al, 2002). Si può dedurre che l'espressione genica non è un indicatore attendibile della espressione proteica e quindi è necessario tenere conto dell'eterogeneità nei meccanismi di sintesi e degradazione, integrando le misure dei livelli di espressione trascritti delle proteine con la misura della velocità con cui vengono sintetizzate.

Per misurare la velocità, sia in vitro che in vivo, si usano metodi basati sulla marcatura delle proteine con traccianti isotopici e sull'utilizzo di modelli precursori-prodotto, permettendo così di misurare il turnover di singole proteine in stato stazionario. Per descrivere il passaggio da uno stato stazionario ad un altro in seguito ad una perturbazione, verranno esaminati alcuni modelli semplici per descrivere la dinamica e i tempi di risposta di alcuni motivi-base di regolazione tra geni e proteine.

³processo di traduzione del gene corrispondente

Capitolo 2

Reti di trascrizione

2.1 Introduzione

La cellula può essere vista come un dispositivo integrato composto da migliaia di tipi di proteine che interagiscono fra di loro. Ogni proteina a sua volta rappresenta un dispositivo molecolare di dimensione nanometrica con specifici compiti eseguiti con grande precisione. In questo elaborato viene usato come esempio il batterio *Escherichia Coli*, un batterio micrometrico, che contiene alcuni milioni di proteine divise in circa 4000 tipi distinti.

Nella sua vita, una cellula incontra diverse situazioni che richiedono specifiche proteine per eseguire vari compiti come la rilevazione di zucchero, caso in cui la cellula produce proteine per trasportarlo al suo interno e per utilizzarlo, oppure quando la cellula è danneggiata, caso in cui essa produce proteine di riparazione. Quindi la cellula monitora in continuazione l'ambiente in cui si trova e calcola la quantità di proteine necessaria per ogni tipo. La funzione che gestisce l'informazione, l'elaborazione e determina il flusso di produzione per ogni proteina è ampiamente effettuata dalla rete di trascrizione.

2.2 Interazione tra cellule e l'ambiente

L'ambiente in cui le cellule vivono è complesso, portando così le cellule a percepire numerosi segnali quali parametri fisici come la temperatura e la pressione osmotica, le molecole biologiche di segnalazione da altre cellule, i nutrienti benefici, le sostanze chimiche dannose. Altre informazioni utili per la cellula riguardano la sua parte interna, cioè la quantità di un deter-

minato metabolita¹ oppure il danneggiamento del DNA, delle membrane o delle proteine all'interno della cellula. In risposta ai vari segnali biologici la cellula produce opportune proteine che interagiscono con l'ambiente interno o esterno. Per raffigurare i vari stati ambientali, la cellula usa come simboli specifiche proteine chiamate fattori di trascrizione.

I fattori di trascrizione sono proteine che legano il DNA in una regione specifica di un promotore o di un enhancer, da dove poi regola la trascrizione. I fattori di trascrizione possono essere attivati o disattivati passando velocemente dallo stato attivo allo stato inattivo con una velocità modulata da uno specifico segnale ambientale (input). Ogni fattore di trascrizione attivo può legare il DNA per regolare la velocità con cui un determinato gene viene letto. I geni vengono letti e trascritti su mRNA, il quale li trasforma in proteine che possono interagire con l'ambiente circostante (fig. 2.1).

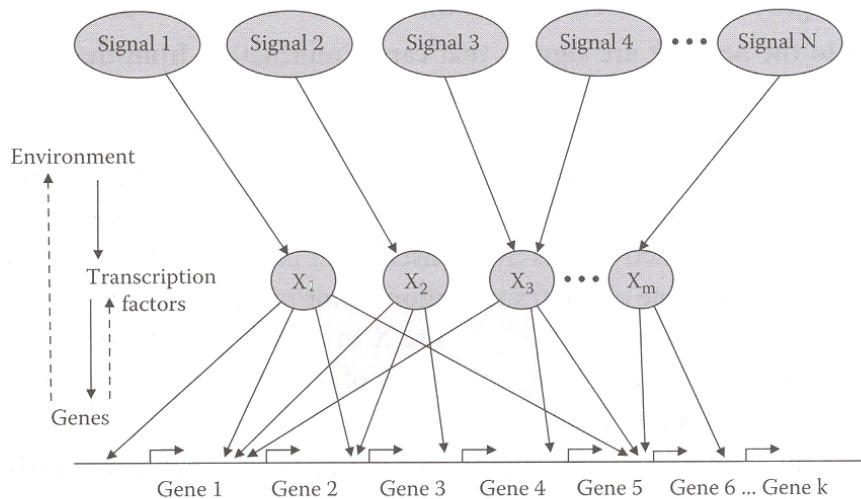


Figura 2.1: Mappa dei segnali, dei fattori di trascrizione e dei geni che essi regolano

L'attività di trascrizione in una cellula può essere considerata una rappresentazione all'interno della cellula dell'ambiente. Per esempio l'E.Coli ha una rappresentazione interna con 300 gradi di libertà (fattori di trascrizione) per regolare il tasso di produzione delle 4000 proteine circa. La rappresen-

¹è il prodotto del processo del metabolismo. Dopo che una sostanza è assimilata, essa subisce un processo di trasformazione che ha il ruolo di rendere la sostanza più assorbibile o eliminabile. Una volta trasformata, una molecola assume il nome di metabolita.

tazione interna con un insieme di fattori di trascrizione è una descrizione molto compatta della miriade di fattori presenti nell'ambiente. Quindi i fattori di trascrizione regolano la produzione dei geni bersaglio per mobilitare la risposta dell'opportuna proteina per un caso specifico.

2.3 Reti di trascrizione

L'interazione tra fattori di trascrizione e geni è descritta dalle reti di trascrizione. Il gene è una parte di DNA la cui sequenza codifica l'informazione necessaria per la produzione di una proteina. La trascrizione di un gene è un processo attraverso il quale l'RNA polimerase (RNAP) produce l'RNA messaggero (mRNA) corrispondente alla sequenza di codifica di quel gene. Successivamente l'mRNA è tradotto in una proteina o in un altro RNA (fig 2.2).

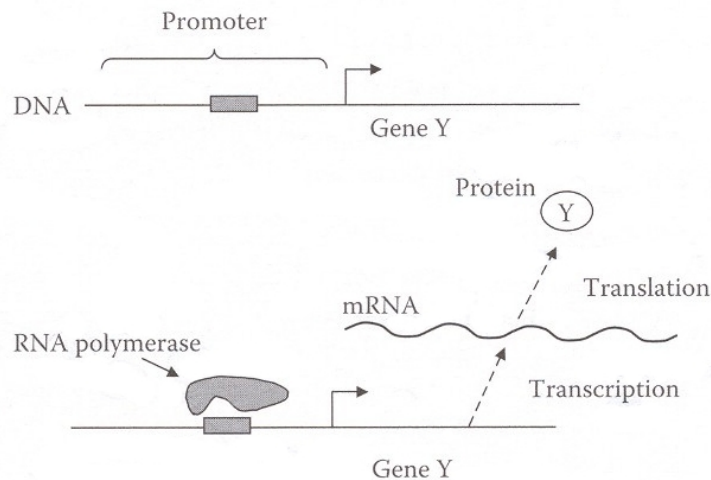


Figura 2.2: Trascrizione di un gene

La velocità con cui il gene viene trascritto (numero di molecole di mRNA prodotte per unità di tempo) è controllata dal promotore, una regione del DNA, che precede il gene, specializzata nella regolazione. L'RNAP lega un determinato sito, cioè una specifica sequenza del DNA, ad un promotore (fig 2.2). La caratteristica di questo sito determina la velocità di trascrizione di questo gene. Anche la sequenza di un sito influisce sulla velocità di

trascrizione, in quanto determina l'affinità chimica dell'RNAP con il sito di produzione.

Considerando che l'RNAP agisce virtualmente su tutti i geni, i cambiamenti all'espressione di geni sono dovuti ai fattori di trascrizione. Ogni fattore di trascrizione modula la velocità di trascrizione di una determinata serie di geni bersaglio.

I fattori di trascrizione influiscono sulla velocità di trascrizione, legando specifici siti dei promotori che agiscono sulla regolazione del gene. Quando sono legati, essi cambiano la probabilità per unità di tempo che l'RNAP ha di attaccare il promotore e produrre una molecola di mRNA.

I fattori di trascrizione possono agire come *attivatori* (fig 2.3), aumentando la velocità di trascrizione del gene, oppure come *repressori* (fig 2.4), riducendo la velocità di trascrizione .

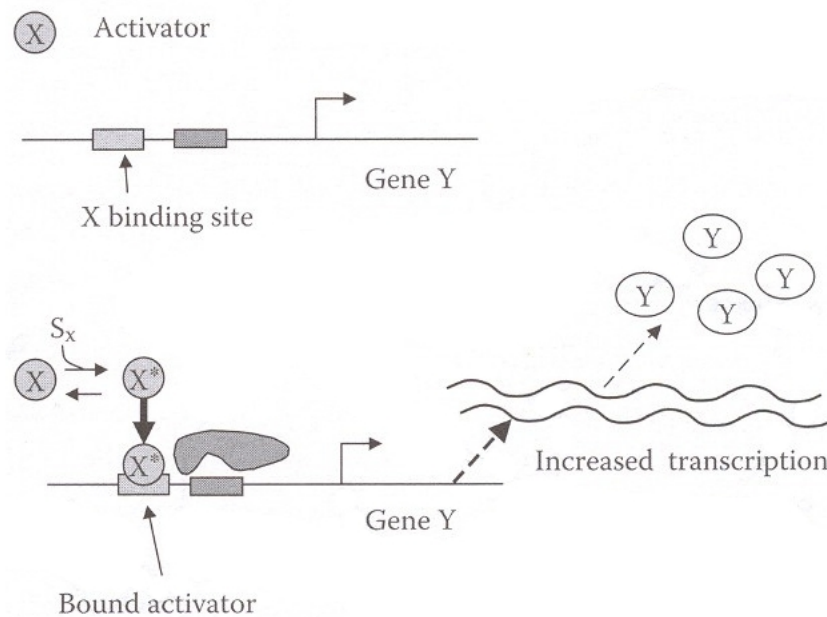


Figura 2.3: Fattore di trascrizione: attivatore

I fattori di trascrizione sono codificati da geni regolati da altri fattori di trascrizione, i quali a loro volta possono essere regolati da altri fattori e così via. Questo insieme di interazioni costituisce una *rete di trascrizione* (fig 2.5).

Una rete di trascrizione descrive tutte le interazioni che regolano le trascrizioni all'interno di una cellula. Nelle reti di trascrizione *i nodi* rappresentano i geni e gli *archi* rappresentano la regolazione trascrittoriale di un gene tramite la

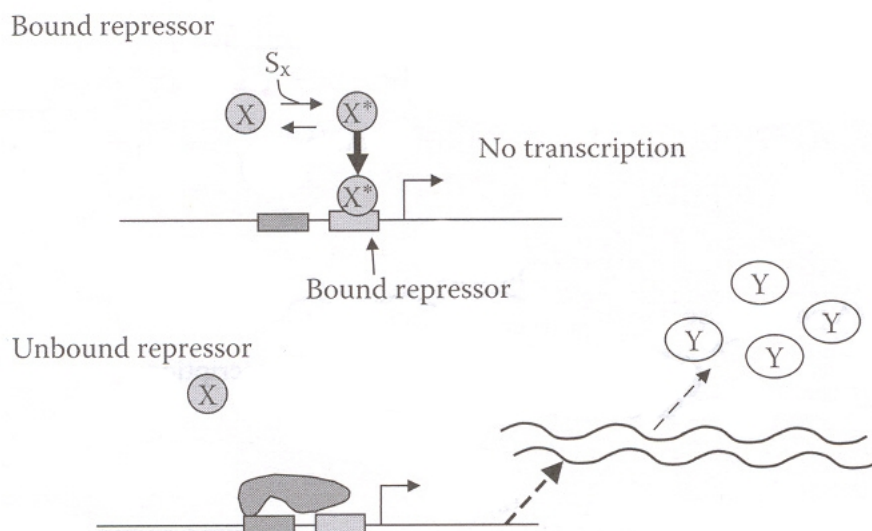


Figura 2.4: Fattore di trascrizione: repressore

produzione di proteine di un altro gene. Un arco diretto $X \rightarrow Y$ descrive come la produzione del gene X, che è un fattore di trascrizione che attacca il promotore del gene Y, controlli la velocità con cui il gene Y è trascritto. Gli ingressi verso la rete di trascrizione sono segnali che portano informazioni dall'ambiente. Ogni segnale è una piccola molecola modificatrice di proteine o una molecola associata che influisce direttamente sull'attività di un fattore di trascrizione. Spesso stimoli esterni attivano nuovi segnali biochimici che portano alla modifica chimica di fattori specifici di trascrizione. In altri sistemi, il segnale può essere molto semplice, come ad esempio la molecola di zucchero che entra nella cellula e lega direttamente il fattore di trascrizione.

Solitamente i segnali causano un cambiamento fisico della forma del fattore di trascrizione, imponendogli di assumere uno stato molecolare attivo. In questo modo, il segnale S_x può causare un rapido cambiamento di X nella sua forma attiva X^* . La forma attiva lega il promotore del gene Y e incrementa il tasso di trascrizione, portando successivamente ad un aumento della produzione della proteina Y. (fig 2.3)

La rete descritta rappresenta una rete dinamica: in seguito all'arrivo di un segnale di ingresso, le attività dei fattori di trascrizione cambiano, portando a delle variazioni nella velocità di produzione delle proteine. Alcune delle proteine sono fattori di trascrizione che attivano altri geni; le proteine

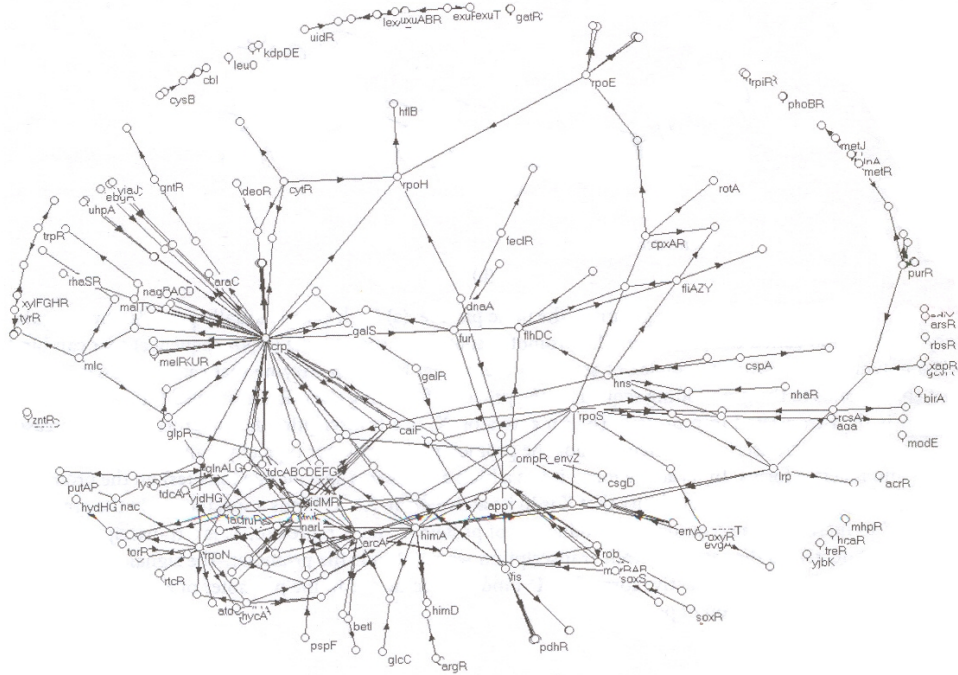


Figura 2.5: Rete di trascrizione

rimanenti non sono fattori di trascrizione, ma si occupano delle varie funzioni delle cellule viventi: come la costruzione delle strutture e la catalizzazione delle reazioni.

I vari fenomeni che si verificano all'interno delle reti di trascrizione presentano diversi ordini di grandezza per la scala dei tempi:

- Solitamente i segnali d'ingresso cambiano l'attività dei fattori di trascrizione in un intervallo di tempo inferiore al secondo;
- Il legame del fattore di trascrizione al proprio sito solitamente raggiunge l'equilibrio nell'arco di un paio di secondi;
- Trascrizione e traduzione del gene bersaglio richiedono un paio di minuti;
- L'accumulazione del prodotto proteico può richiedere da alcuni minuti fino a più ore.

Si può così osservare che ci sono ordini di grandezza temporali diversi tra il segnale d'ingresso e l'accumulazione del prodotto proteico, cosa che si può osservare anche dalla tabella 2.1 per quanto riguarda *E. Coli*. In questo modo

Legame di una piccola molecola (un segnale) a un fattore di trascrizione	~ 1 msec
Legame di un fattore di trascrizione attivo al suo sito DNA	~ 1 sec
Trascrizione e traduzione di un gene	~ 5 min
Tempo necessario per raggiungere il 50% della soglia di concentrazione di una proteina tradotta	~ 1 ore

Tabella 2.1: Scala dei tempi nella rete di trascrizione dell'E. coli

si possono assumere i livelli di attività dei fattori di trascrizione come costanti nelle equazioni che descrivono la dinamica della rete sulla scala dei tempi del cambiamento dei livelli proteici, molto più lenta di quella di attività dei fattori di trascrizione. In aggiunta alle reti di trascrizione, le cellule contengono tante altre reti di interazione come le reti con segnali di trasduzione composte da proteine interagenti fra di loro. Queste reti tipicamente sono molto più veloci delle reti di trascrizione e possono essere approssimate anche queste in stato stazionario rispetto alle reti di trascrizione, semplificando di molto l'analisi.

C'è una ricca varietà di meccanismi tramite i quali i fattori di trascrizione regolano i geni, e qui la biologia mostra la sua immensa complessità. I fattori di trascrizione mostrano i modi ingegnosi attraverso i quali legano il DNA dei siti strategicamente posizionati. Quando sono legati, essi si bloccano o si reclutano reciprocamente assieme all'RNAP per controllare il flusso di produzione dell'mRNA. Tuttavia, al livello di dinamica delle reti di trascrizione e grazie alla lenta scala dei tempi su cui operano, si possono trattare tutti questi meccanismi con una semplice descrizione matematica unificante e piuttosto semplice.

Un'altra notevole proprietà delle reti di trascrizione è la modularità dei loro componenti. Si può prendere il DNA di un gene di un organismo ed esprimerlo all'interno di un organismo diverso. Per esempio si può prendere la regione di codifica del DNA per la proteina GFP² dal genoma di una medusa e introdurlo in un batterio. Come risultato il batterio produce GFP, diventando così verde. Inoltre può essere aggiunta la regolazione inserendo una regione promotore. Per esempio il controllo del gene GFP nel batterio può essere attivato aggiungendo davanti al gene un frammento di DNA da un promotore di un gene batterico diverso, come ad esempio quello che controlla la presenza di zucchero nella cellula. Questo permette all'E. Coli di diventare verde solo in presenza di zucchero al suo interno. Promotori e geni sono generalmente interscambiabili e per questo motivo la GFP è diventata negli

²La green fluorescent protein è proteina espressa nella medusa *Aequorea victoria*.

ultimi decenni un diffuso strumento per esperimenti e tecniche di biologia molecolare.

Tutti i componenti modulabili fanno sì che la rete di trascrizione sia molto flessibile durante la sua evoluzione e le permettono di incorporare prontamente nuovi geni e nuove regolazioni. Infatti le reti di trascrizione possono evolvere rapidamente, non a caso sembra che gli archi nella rete di trascrizione si evolvano più velocemente della regione di codifica dei geni.

Per esempio: animali imparentati, come l'uomo e il topo, hanno geni molto somiglianti, ma la regolazione della trascrizione di questi geni, che governano quando e quanto siano prodotte le proteine, è evidentemente molto diversa.

In altre parole, tante differenze tra le specie animali sembra siano dovute alla differenza negli archi delle reti di trascrizione piuttosto che alle differenze nei loro geni.

2.4 Gli archi nelle reti di trascrizione

Ogni arco in una rete di trascrizione corrisponde ad un'interazione in cui il fattore di trascrizione controlla direttamente la velocità di trascrizione di un gene. Queste interazioni possono essere di 2 tipi: *attivazione* (controllo positivo) oppure *repressione* (controllo negativo). L'attivazione avviene quando il fattore di trascrizione incrementa il tasso di trascrizione nella fase di legame al promotore (fig 2.3). In questo caso l'arco ha il segno +. La repressione avviene quando il fattore di trascrizione riduce il tasso di trascrizione nella fase di legame al promotore (fig 2.4) e l'arco relativo ha il segno -.

Tipicamente i fattori di trascrizione agiscono principalmente o come attivatori o come repressori, ma possono agire anche in entrambi i modi. In altre parole i segni degli archi uscenti da un determinato nodo, che rappresentano così i geni regolati da quel nodo, sono altamente correlati fra di loro. Alcuni nodi mandano principalmente archi con segno negativo, rappresentando dei repressori, in altri nodi manda fuori principalmente archi con segno positivo, che sono attivatori. Tuttavia la maggior parte degli attivatori che regolano tanti geni agiscono anche come repressori per alcuni dei geni target. Lo stesso ragionamento può essere applicato anche per i repressori, che possono agire positivamente su una parte dei geni che controllano.

Così i fattori di trascrizione tendono ad impiegare un unico modo di regolazione per la maggioranza dei loro geni bersaglio. In contrasto con ciò, i segni sugli archi che entrano in un nodo, che rappresentano le interazioni di trascrizione che regolano il gene, sono meno correlati. Tanti geni, controllati

da fattori di trascrizione multipli, presentano ingressi attivatori da alcuni fattori e repressori da altri.

In breve i segni che escono da un arco (gli archi puntano fuori da un dato nodo) sono piuttosto correlati, mentre i segni degli archi entranti (che puntano al nodo) non lo sono.

Gli archi non hanno solo i segni + o -, ma possono avere anche numeri che corrispondono all'intensità dell'interazione. L'intensità dell'effetto di un fattore di trascrizione sulla velocità di trascrizione del suo gene bersaglio viene descritta da una funzione. Prendiamo in considerazione il livello di espressione della proteina Y controllato da un unico fattore di trascrizione X . Esso viene rappresentato nella rete da $X \rightarrow Y$, con il numero di molecole della proteina Y prodotte per unità di tempo funzione della concentrazione di X nella sua forma attiva X^* :

$$\text{livello di espressione } Y = f(X^*)$$

2.4.1 Funzione di Hill

Tipicamente la funzione $f(X^*)$ è monotona, crescente quando X è un attivatore (fig 2.6) e decrescente quando X è un repressore (fig 2.7). Una funzione molto utile che descrive il livello di espressione di molti geni reali è la funzione di Hill. Questa funzione può essere ricavata dalle considerazioni sul legame all'equilibrio del fattore di trascrizione al suo sito promotore. La *funzione di Hill* per un attivatore è una curva che sale da zero e si avvicina al valore di regime:

$$f(X^*) = \frac{\beta X^{*n}}{K^n + X^{*n}} = \beta \frac{(X^*/K)^n}{1 + (X^*/K)^n} \quad (2.1)$$

La funzione ha 3 parametri K , β e n . Il primo parametro K è il *coefficiente di attivazione* e misura la concentrazione di X attiva necessaria affinché l'espressione sia attiva. Dall'equazione si può osservare che per raggiungere metà della massima espressione bisogna avere $X^* = K$ (fig 2.6). Il valore di K è collegato all'affinità chimica tra X e il suo sito sul promotore. Il secondo parametro nella funzione ingresso è il massimo livello di produzione β . La massima produzione viene raggiunta ad elevate concentrazioni di attivatore $X^* > K$, in quanto X^* lega il promotore con maggiore probabilità e stimola l'RNAP a produrre molte più unità di mRNA per unità di tempo. Il coefficiente n della funzione di Hill gestisce la ripidità della funzione: più grande è n e più la funzione assomiglia ad una funzione soglia (fig 2.6).

Così come tante funzioni in ambito biologico, la funzione di Hill raggiunge un valore soglia per valori elevati di X^* , piuttosto che crescere infinitamente. La soglia della funzione di Hill è principalmente dovuta al fatto che

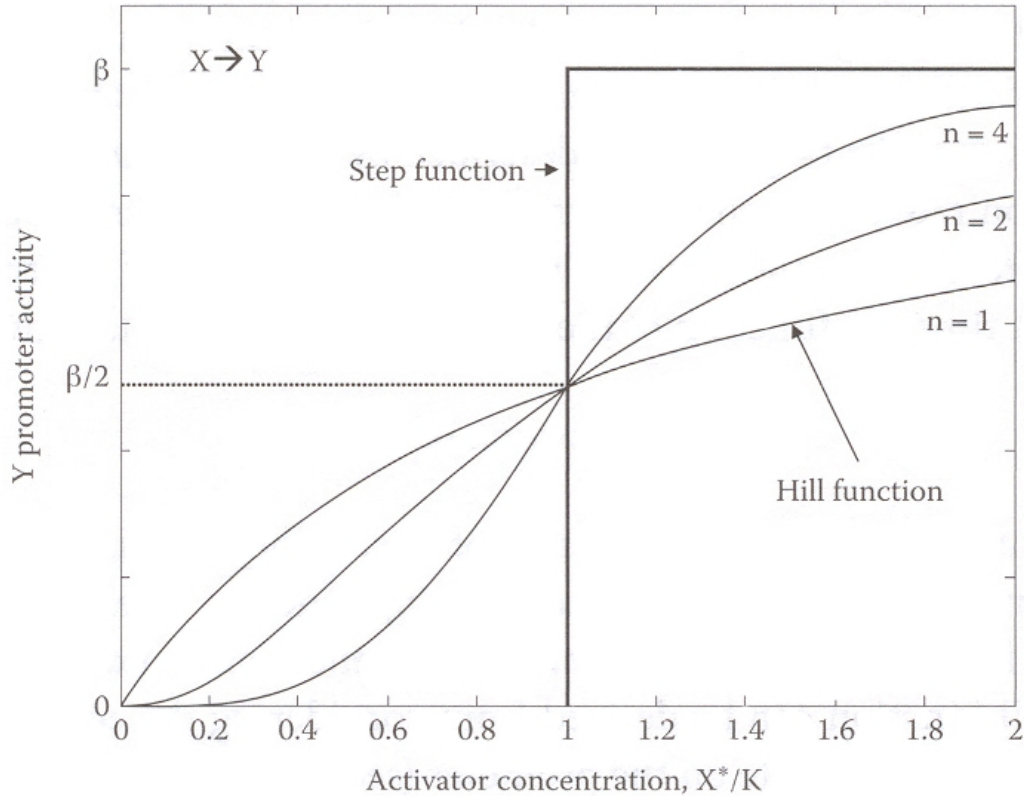


Figura 2.6: Funzione di Hill per un fattore di trascrizione attivatore

la probabilità che un attivatore leghi il suo promotore non può superare 1, indistintamente da quanto grande sia la concentrazione di X^* .

Per un repressore, la funzione di Hill è una curva decrescente la cui forma dipende dai 3 parametri citati precedentemente:

$$f(X^*) = \frac{\beta K^n}{K^n + X^{*n}} = \frac{\beta}{1 + (X^*/K)^n} \quad (2.2)$$

Siccome un repressore permette la trascrizione di un gene quando non è legato al promotore, questa funzione può essere vista come la probabilità che X^* non leghi il promotore. Il massimo flusso di produzione β è ottenuto quando il repressore non lega il promotore per niente (fig 2.4), ossia quando $X^* = 0$.

Metà del valore massimo di repressione viene raggiunto quando l'attività del repressore è uguale a K (coefficiente di repressione). Il coefficiente n

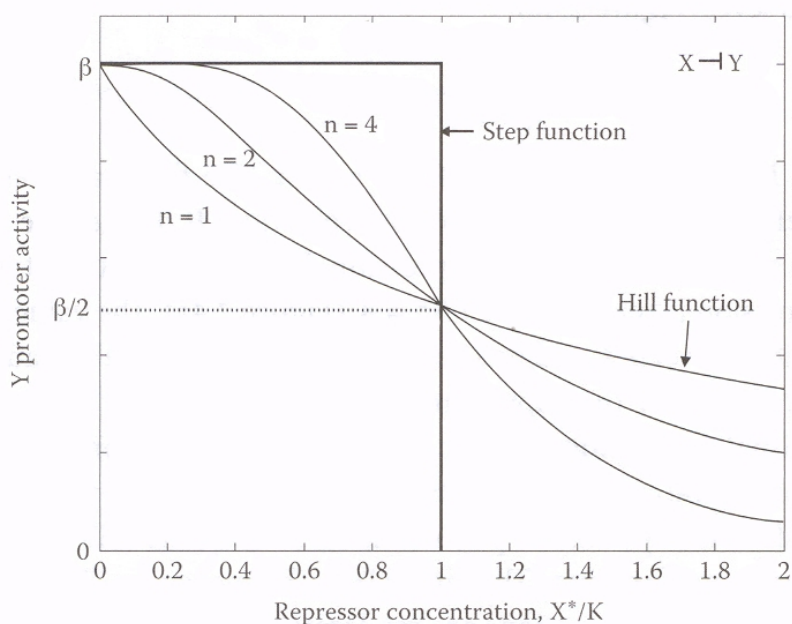


Figura 2.7: Funzione di Hill per un fattore di trascrizione repressore

determina anche in questo caso la ripidità della funzione (fig 2.7).

Da quest'analisi ne risulta che ogni arco della rete può avere 3 parametri: β , K ed n . Questi parametri possono essere regolati velocemente durante l'evoluzione della cellula. Per esempio, K può essere cambiato da mutazioni che alterano la sequenza di DNA del sito promotore del gene Y legato da X . Anche il cambiamento di una singola lettera nella sequenza di DNA cambia K . Il parametro K può cambiare anche se viene cambiata la posizione del sito di legame, così come da scambi nelle sequenze fuori da quella del sito di legame (questi effetti non sono pienamente compresi tuttora) (U.Alon, 2007). Analogamente, il massimo livello di produzione β può essere modificato da mutazioni nel sito di legame del RNAP.

Gli esperimenti di laboratorio mostrano che una volta che un batterio viene posto in un nuovo ambiente, questo può modificare questi parametri con qualche centinaio di generazioni, in modo da raggiungere il livello di espressione ottimale. In altre parole, questi numeri sono in continuo cambiamento e possono ereditare tanti cambiamenti se l'ambiente cambia.

2.4.2 Funzioni logiche

Le funzioni di Hill sono utili per modelli dettagliati. Per un'analisi qualitativa, è utile usare funzioni più semplici che mostrano intuitivamente il funzionamento essenziale di queste funzioni che descrivono il livello di espressione della proteina Y controllato da un unico fattore di trascrizione X . L'essenza di queste funzioni è la transizione tra livello basso e alto con soglia K . Si possono approssimare queste funzioni nelle reti di trascrizione con con porte logiche (Glass and Kauffman, 1973; Thieffry & Thomas, 1998). Con questa approssimazione, la produzione dei geni può essere inattiva ($f(X^*) = 0$), oppure attiva, al valore massimo: $f(X^*) = \beta$. La soglia per l'attivazione è K . Quindi, queste funzioni logiche sono delle approssimazioni in forma di gradino per le funzioni di Hill, le quali hanno un andamento più sinuoso. Per gli attivatori, la funzione logica può essere descritta utilizzando la funzione gradino θ che presenta un salto quando X^* supera la soglia K .

$$f(X^*) = \beta\theta(X^* > K) \quad \text{approssimazione logica per attivatori} \quad (2.3)$$

dove θ è uguale a 0 o a 1 in base al valore logico dell'argomento. L'approssimazione logica equivale ad una funzione di Hill con pendenza elevata, cioè $n \rightarrow \infty$.

Analogamente, per i repressori, l'approssimazione è una funzione gradino decrescente:

$$f(X^*) = \beta \cdot \theta(X^* < K) \quad \text{approssimazione logica per un repressore} \quad (2.4)$$

2.4.3 Funzioni multidimensionali che regolano i geni con vari input

Come appena visto, le funzioni di Hill e le funzioni logiche possono descrivere l'input da un singolo fattore di trascrizione. Tuttavia tanti geni sono regolati da fattori di trascrizione multipli, ossia tanti nodi nella rete hanno due o più archi entranti. L'attività dei loro promotori è quindi una funzione multidimensionale con diversi fattori di trascrizione come input (Yuh et al., 1998; Pipel et al., 2001; Buchler et al., 2003, Setty et al., 2003). In questi casi è utile approssimare le funzioni multi-dimensionali con funzioni logiche come nel caso delle funzioni unidimensionali. Per esempio consideriamo i geni regolati da 2 attivatori: i geni richiedono il legame di entrambe le proteine attivatrici al promotore per avere una significativa espressione. Questa funzione è simile alla porta logica AND.

$$f(X^*, Y^*) = \beta \cdot \theta(X^*)\theta(Y^*) \sim X^* \text{ AND } Y^* \quad (2.5)$$

Per altri geni, è sufficiente il legame di uno dei 2 geni, e questo è il caso della porta OR: $f(X^*, Y^*) = \beta \cdot \theta(X^* > K_x \text{ OR } Y^* > K_y)$. Però non tutti i geni hanno una funzione di ingresso di tipo booleano: per esempio certi geni hanno una funzione di tipo somma in cui gli ingressi sono di tipo additivo (Kalir & Alan, 2004): $f(X^*, Y^*) = \beta_x X^* + \beta_y Y^*$. Ci possono essere anche altri tipi di funzioni simili ad una porta OR, oppure ad una funzione di tipo somma in cui gli ingressi sono di tipo additivo o addirittura ad una funzione con tante parti costanti a tratti e soglie diverse come nel sistema del promotore lac³ di E.Coli.

La forma di queste funzioni può essere rapidamente cambiata da mutazioni al promotore del gene regolato. Per esempio, la funzione del lac può essere modificata per assomigliare ad una porta AND oppure OR con poche mutazioni nel promotore lac (Mayo et al. 2006). Pare che l'esatta forma della funzione ingresso per ogni gene sia in continua modifica e selezione durante l'evoluzione della cellula.

2.5 La regolazione di un gene

Le reti di trascrizione descrivono la regolazione della trascrizione dei geni. Ogni nodo rappresenta un gene. Gli archi sono identificati da $X \rightarrow Y$ e descrivono come il gene X codifica il fattore di trascrizione (proteina) che legga il promotore del gene Y e controlla la sua velocità di trascrizione.

Così le proteine codificate dal gene X modificano il flusso di produzione della proteina codificata dal gene Y . La proteina Y , a sua volta, può essere fattore di trascrizione che cambia la velocità di produzione di Z , e così via, formando una rete interattiva. La maggior parte dei nodi nella rete rappresentano geni che codificano proteine che non sono fattori di trascrizione. Queste proteine si occupano delle varie funzioni biologiche all'interno della cellula.

Gli ingressi nella rete sono segnali che ricavano informazioni dall'ambiente e cambiano l'attività di specifici fattori di trascrizione.

I fattori di trascrizione attivi legano determinati siti di DNA nei promotori dei loro geni bersaglio per controllare la velocità di trascrizione. Questo è descritto quantitativamente dalle funzioni ingresso: la velocità di produzione del gene prodotto Y è funzione della concentrazione di fattori di trascrizione attivi X^* . I geni regolati da multipli fattori di trascrizione sono descritti da

³utilizzo del lattosio dell'E. coli

funzioni multi-dimensionali. Le funzioni sono spesso ripide e possono essere approssimate dalla funzione di Hill o da porte logiche.

Ogni arco e ogni funzione ingresso sono sottoposte alla selezione naturale. Un arco inutile sarà velocemente perso con le mutazioni genetiche. È sufficiente il cambio di una o più lettere nella sequenza di DNA nel sito di legame di X nel promotore di Y per cancellare l'arco $X \rightarrow Y$.

Per analizzare la dinamica di un unico arco nella rete di trascrizione viene preso in considerazione un gene regolato da un unico regolatore, senza input addizionali. La sua trascrizione è descritta nella rete da $X \rightarrow Y$ (il fattore di trascrizione X regola il gene Y). Una volta che X diventa attivo per mezzo di un segnale, la concentrazione di Y inizia a cambiare.

In assenza del suo segnale di ingresso, X è inattivo e Y non è prodotta (fig 2.3). All'arrivo del segnale S_x , X transita rapidamente nella sua forma attiva X^* e lega il promotore del gene Y . Il gene Y inizia ad essere trascritto, e l'mRNA è tradotto, avendo così come risultato l'accumulazione della proteina di Y .

La concentrazione di una proteina in un cellula è data dal bilanciamento tra la sua produzione (tramite la trascrizione di un gene, seguita dalla traduzione di mRNA codificante la stessa), la degradazione dovuta alla sua demolizione da parte di proteine specializzate all'interno della cellula e la riduzione di concentrazione conseguente all'aumento del volume di una cellula durante la sua fase di crescita.

Considerando $Y(t)$ l'espressione di una determinata proteina all'interno di una cellula, $P(t)$ il suo flusso di produzione ed $U(t)$ il flusso che comprende la degradazione e la diluizione, si può esprimere una relazione matematica:

$$\frac{dY(t)}{dt} = P(t) - U(t) \quad (2.6)$$

Buona parte delle proteine intracellulari si trovano in uno stato stazionario definito da valori costanti di concentrazione Y^{ss} , produzione P^{ss} e degradazione o diluizione U^{ss} , in quanto sono costantemente degradate e risintetizzate. Per analizzare uno stato stazionario, indichiamo con β la costante di produzione e con α il rapporto tra i valori di stato stazionario di utilizzazione e concentrazione ($\alpha = \frac{U^{ss}}{Y^{ss}}$), che misura la velocità di turnover. L'equazione precedente diventa

$$\frac{dY(t)}{dt} = 0 = \beta - \alpha Y^{ss} \quad (2.7)$$

Nello stato di equilibrio, Y raggiunge una concentrazione costante Y^{ss} . La concentrazione Y^{ss} può essere calcolata risolvendo l'equazione precedente. Questo mostra che la concentrazione Y^{ss} è il rapporto fra produzione e utilizzazione: $Y^{ss} = \frac{\beta}{\alpha}$. Si può osservare che mantenendo costante la produzione β

e aumentando il turnover α , diminuisce la concentrazione Y^{ss} della proteina. In più, un valore elevato di α richiede un grande flusso di produzione bilanciato da un elevato flusso di degradazione. Questo ciclo potrebbe sembrare inutile, però una produzione elevata di proteine ottenuta assemblando una serie di aminoacidi, bilanciato dalla loro scomposizione permette di preservare la qualità della proteina e quindi l'integrità della funzione biologica da essa svolta.

Di seguito vengono analizzati due casi estremi: uno in cui la produzione di una proteina cessa e l'altro in cui la proteina è prodotta per la prima volta all'interno della cellula.

Nel primo caso la concentrazione e l'utilizzazione della proteina scendono nel tempo. Supponendo che il loro rapporto si mantiene costante nel tempo pari al valore iniziale α , l'equazione iniziale diventa:

$$\frac{dY(t)}{dt} = -U(t) = -\alpha Y(t) \quad Y(0) = Y^{ss} = \frac{\beta}{\alpha} \quad (2.8)$$

la cui soluzione è un andamento esponenziale decrescente che scende a zero (fig. 2.8):

$$Y(t) = Y^{ss} e^{-\alpha t} = \frac{\beta}{\alpha} e^{-\alpha t} \quad (2.9)$$

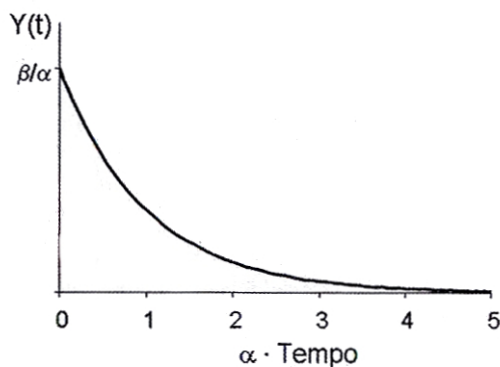


Figura 2.8: Andamento esponenziale decrescente della concentrazione di una proteina in risposta all'assenza di un segnale di ingresso

Per analizzare quanto velocemente avviene il decadimento del livello di Y si analizza il *tempo di risposta*. Il tempo di risposta $T_{1/2}$, è definito come il tempo necessario affinché si raggiunga la metà tra il livello iniziale e il

livello finale nel processo dinamico. Quindi per il processo di decadimento di Y , il tempo di risposta si calcola risolvendo l'equazione $Y(t) = \frac{Y^{ss}}{2}$, per cui abbiamo una dipendenza inversa da α :

$$T_{1/2} = \frac{\log(2)}{\alpha} \quad (2.10)$$

Bisogna osservare come il flusso di utilizzazione α influisca direttamente sul tempo di risposta: flussi di utilizzazione veloci determinano cambiamenti veloci di concentrazione. Il flusso di produzione β invece influisce sul livello di stato stazionario, ma non sul tempo di risposta. I benefici di un flusso di utilizzazione elevato sono i tempi di risposta ridotti in presenza di una perturbazione subita, permettendo di mantenere la funzione biologica inalterata.

Consideriamo ora il caso contrario in cui inizialmente la cellula non è stimolata per la produzione di Y ($Y = 0$) e in seguito al segnale S_x inizia la sua produzione. La produzione della proteina sale istantaneamente da 0 al valore β , avendo una crescita esponenziale dell'espressione proteica che si stabilizza al suo valore di regime (fig 2.9):

$$Y(t) = Y^{ss} (1 - e^{-\alpha t}) = \frac{\beta}{\alpha} (1 - e^{-\alpha t}) \quad (2.11)$$

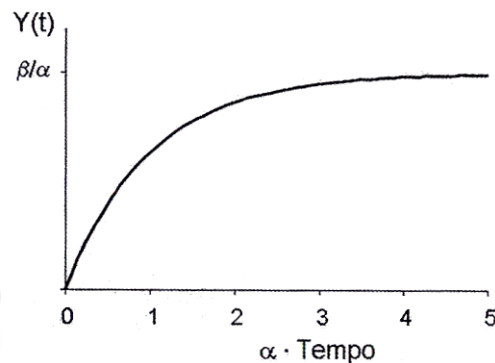


Figura 2.9: Andamento esponenziale della concentrazione di una proteina in risposta ad una salita a gradino

Per $t \ll 1$ possiamo usare lo sviluppo in serie di Taylor per trovare un' approssimazione lineare dell'accumulazione di Y :

$$Y \sim \beta t \quad \text{per tempi inferiori a } 1 \quad t \ll 1 \quad (2.12)$$

Questo ha senso in quanto la concentrazione di Y aumenta per tempi inferiori al secondo proporzionalmente al suo flusso di produzione. Successivamente, il livello di Y aumenta esponenzialmente finché Y converge asintoticamente al livello di regime Y^{ss} .

Il tempo di risposta è il tempo necessario affinché $Y(t)$ raggiunga il valore $\frac{Y^{ss}}{2}$. Questo tempo può essere ricavato risolvendo l'equazione $Y(t) = \frac{Y^{ss}}{2}$ nell'incognita t . Ne segue che anche in questo caso il tempo di risposta è lo stesso del caso precedente e dipende dal flusso di utilizzazione della proteina. A flussi di utilizzazione maggiori corrispondono brevi tempi di risposta per il cambiamento della concentrazione.

2.5.1 Tempo di risposta di una proteina stabile

Un tipo particolare di proteine, denominate proteine stabile, non viene degradato nella cellula ($\alpha_{deg} = 0$). La loro produzione è bilanciata dalla sola diluizione legata all'aumento del volume della cellula durante la fase di crescita ($\alpha = \alpha_{dil}$). Per queste proteine, il tempo di risposta coincide con il tempo di replicazione cellulare, quindi $T_{1/2}$ corrisponde alla durata del ciclo cellulare, che può variare da 30 minuti a svariate ore e in alcuni eucarioti i tempi sono addirittura più lunghi.

Per osservare questo, immaginiamo che una cellula produce una proteina e di colpo la produzione si ferma ($\beta = 0$). La cellula cresce e quando raddoppia il suo volume si divide in 2 cellule. Così, dopo il tempo di generazione di una cellula (τ), la concentrazione della cellula dimezza e quindi:

$$T_{1/2} = \frac{\log(2)}{\alpha_{dil}} = \tau \quad \text{tempo di risposta di generazione di una cellula} \quad (2.13)$$

Ci si aspetta che le reti di trascrizione abbiano tempi di risposta a segnali come i nutrienti o stress esterni dello stesso ordine di quelli della produzione. Per proteine stabili, i tempi di risposta, come visto in precedenza, sono i tempi di produzione di una cellula; tempi che possono essere un fattore limitante mettendo un vincolo per la progettazione efficiente delle di trascrizione dei geni.

Riassumendo, i tempi di risposta della regolazione di un singolo gene dipendono dai flussi di utilizzazione della proteina. Nel capitolo successivo verranno analizzati delle semplici reti di trascrizione, che possono aiutare a velocizzare i tempi di risposta.

Autoregolazione

3.1 Introduzione

Nel precedente capitolo è stata analizzata la dinamica di base di una sola interazione nelle reti di trascrizione. In seguito, l'analisi è ampliata al funzionamento di una rete di trascrizione reale fatta da tante interazioni (Fig. 3.2). Come esempio, viene usata una rete di trascrizione dell'E. coli che include circa il 20% dei geni dell'organismo (Shen-Orr et al., 2002). Questa rete è molto complessa e l'obiettivo è quello di definire la connessione tra pattern noti che servono alla costruzione della rete. In seguito si definisce un modo per determinare i pattern denominati motivi in una rete complessa, si esaminano i motivi più semplici in una rete di trascrizione, cioè l'autoregolazione negativa e si mostra che questo motivo ha una funzione utile: velocizzare i tempi di risposta delle interazioni nella fase di trascrizione e stabilizzarle.

3.2 Reti reali e reti random

La rete di trascrizione dell'E. coli contiene numerosi pattern di nodi e archi. La selezione dei pattern più importanti è effettuata in base alla significatività statistica¹. Per definire la significatività statistica, bisogna confrontare la rete di trascrizione con un insieme di reti random; reti che hanno le stesse caratteristiche delle reti reali (per esempio hanno lo stesso numero di nodi e archi come le reti reali) che però presentano connessioni aleatorie fra nodi e archi. I pattern che si riscontrano molto più spesso nelle reti reali piuttosto che il quelle random prendono il nome di motivi (Milo et al., 2002; Shen-Orr et al., 2002).

¹la possibilità rilevante che compaia un determinato valore

Se un motivo compare in una rete di trascrizione più spesso che in una rete random, vuol dire che è stato selezionato perché offre dei vantaggi all'organismo. Se il motivo non avrebbe offerto un vantaggio, sarebbe stato cancellato e sarebbe comparso con la stessa probabilità che ha nelle reti random.

Per trovare i motivi, dobbiamo confrontare una rete reale con un insieme di reti random. Per la descrizione dell'insieme delle reti random si usa un modello semplice introdotto da Erdos e Renyi (Erdos e Renyi, 1959; Bollobus, 1985). Questo modello permette di semplificare i conti offrendo le stesse risposte qualitative di modelli di reti random più complessi.

Una rete di trascrizione reale ha N nodi e M archi. Per confrontarla con il modello Erdos-Renyi (ER), bisogna costruire una rete random con lo stesso numero di nodi e archi. Nelle reti random definite dal modello ER, gli archi diretti tra le coppie di nodi sono assegnati in modo aleatorio.

Siccome ci sono N nodi risultano $\frac{N(N-1)}{2}$ possibili coppie di nodi che possono essere collegate da archi. Ogni arco può puntare in una delle due direzioni, per un totale di $N(N-1)$ possibili combinazioni. In più, un arco può iniziare e finire nello stesso nodo, formando così un cappio (in tutto N). Quindi le possibili combinazioni sono

$$N(N-1) + N = N^2 \quad (3.1)$$

Nel modello ER, gli M archi sono posizionati random in N^2 possibili posizioni, quindi ogni possibile posizione occupata dall'arco ha una probabilità $p = \frac{M}{N^2}$. La figura 3.1 confronta una piccola rete alla rete random ER corrispondente con stesso numero di nodi e archi.

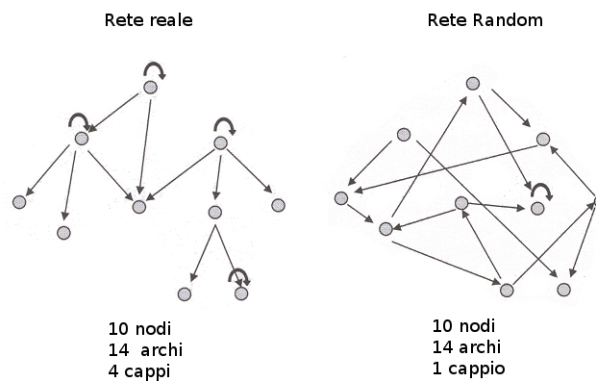


Figura 3.1: Confronto tra una rete reale e una rete random

3.3 Autoregolazione: motivi

La regolazione di un gene da parte del prodotto del gene stesso è conosciuto come controllo autogeno, o autoregolazione e nella rete è individuata da un cappio, un arco che parte da un nodo e arriva nello stesso nodo. Per vedere se l'autoregolazione è un motivo bisogna verificare se la sua ricorrenza è più frequente in una rete reale piuttosto che in una random.

Per decidere ciò, viene usata come rete reale la rete dell'E. coli della fig. 3.2 che ha 40 cappi, di cui 34 sono autoregolazione negativa. In più è necessario calcolare la probabilità di avere K cappi in una rete random ER. Per formare un cappio, un arco ha la necessità che il nodo di origine e quello di destinazione siano lo stesso, uno solo degli N possibili nodi. Così abbiamo la probabilità: $p_{auto} = \frac{1}{N}$

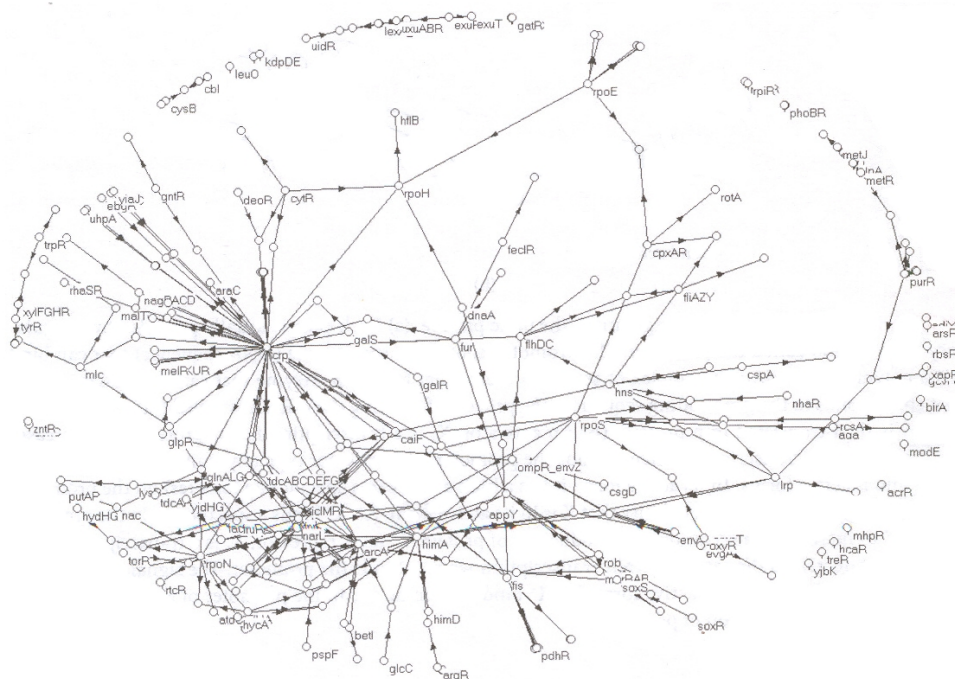


Figura 3.2: Rete di trascrizione

Poiché ci sono M archi posizionati random per formare la rete random, la probabilità di avere K cappi è circa binomiale (l'evento descritto è: lanciando una moneta M volte voglio K teste):

$$P(K) = \binom{M}{K} p_{auto}^K (1 - p_{auto})^{M-K} \quad (3.2)$$

Il valore medio di cappi è uguale al numero di archi M per la probabilità che un arco sia cappio (esattamente il numero di teste è il numero di volte che la moneta è lanciata moltiplicato per la probabilità che esca testa):

$$\langle N_{auto} \rangle_{rand} \sim M \cdot p_{auto} \sim \frac{M}{N} \quad (3.3)$$

con deviazione standard che è circa la radice della media, nuovamente simile al lancio di una moneta con probabilità di testa p_{auto} molto piccola, e approssima un processo di Poisson:

$$\sigma_{auto} = \sqrt{\frac{M}{N}} \quad (3.4)$$

Nella rete di trascrizione dell'E. coli in fig 3.2, il numero di nodi è $N = 424$ e quello degli archi $M = 519$. Una rete ER corrispondente con gli stessi N ed M , dovrebbe avere un solo cappio con uno scarto di ± 1 circa:

$$\langle N_{auto} \rangle_{rand} \sim \frac{M}{N} \sim 1.2 \quad \sigma_{rand} = \sqrt{1.2} \sim 1.1 \quad (3.5)$$

In contrasto con ciò, una rete reale ha 40 cappi, superando di gran lunga il limite della deviazione standard della rete random, descrivendo come siano molto più frequenti di quella random. Così, i cappi, e in particolare quelli autoregolati negativi, sono dei motivi.

3.4 Vantaggi dell'autoregolazione

Ci si può chiedere perché l'autoregolazione negativa è un motivo e quali vantaggi offre. Per rispondere a ciò, confronteremo un gene autoregolato negativamente con un gene senza autoregolazione (fig 3.3). Il criterio di confronto sarà il tempo di risposta del sistema.

L'autoregolazione negativa si verifica quando un fattore di trascrizione X reprime la propria trascrizione (fig 3.3b). Questa autorepressione si verifica quando X lega il proprio promotore per bloccare la produzione di $mRNA$. Come risultato, maggiore è la concentrazione di X , e minore è il suo flusso di produzione. Come visto nel capitolo precedente, la dinamica di X è descritta dal suo flusso di produzione $f(X)$ e dal suo flusso di utilizzazione:

$$\frac{dX}{dt} = f(X) - \alpha X \quad (3.6)$$

dove $f(X)$ è una funzione decrescente di X . Come menzionato nel capitolo 2, una buona approssimazione per tanti promotori è la funzione decrescente di

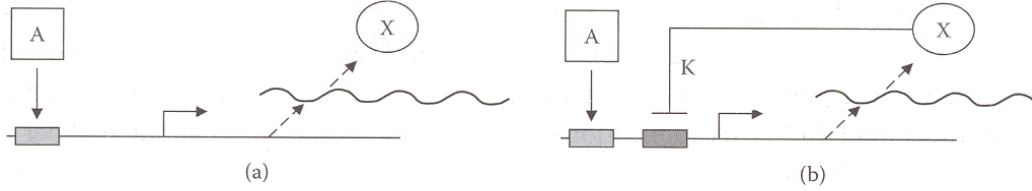


Figura 3.3: Confronto tra gene senza autoregolazione (a) e gene con autoregolazione negativa (b)

Hill :

$$f(x) = \frac{\beta}{1 + \left(\frac{x}{K}\right)^n} \quad (3.7)$$

Quando X è molto minore del coefficiente di repressione K , il promotore è libero e il suo flusso di produzione raggiunge il valore massimo β . Dall'altro lato, quando il repressore X ha una concentrazione elevata, non c'è alcuna trascrizione, $f(X) \sim 0$. Per descrivere l'andamento della funzione in maniera più intuitiva, usiamo l'approssimazione ad una funzione logica, dove la produzione è nulla se $X > K$, e la produzione è massimale, $f(X) = \beta$, quando $X < K$. Questo è stato descritto nel capitolo 2 usando la funzione a soglia θ :

$$f(X) = \beta \cdot \theta(X < K) \quad (3.8)$$

Per studiare i tempi di risposta, consideriamo il caso in cui X inizialmente non c'è e la sua produzione inizia per $t = 0$. Per tempi più brevi, mentre la concentrazione di X è bassa, il promotore non è represso e la produzione avviene alla massima velocità β , come descritto dall'equazione produzione-degradazione:

$$\frac{dX}{dt} = \beta - \alpha \cdot X \quad \text{mentre } X < K \quad (3.9)$$

Questo risultato si avvicina molto al livello massimo dello steady state descritto nel capitolo precedente. Per tempi brevi, infatti, la degradazione può essere ignorata ($X \ll \beta$), trovando così un'accumulazione lineare di X in funzione del tempo:

$$X(t) \sim \beta t \quad \text{con } X < K \text{ e } X \ll \frac{\beta}{\alpha} \quad (3.10)$$

Tuttavia, la produzione si ferma quando il livello X raggiunge la soglia di autorepressione, $X = K$ (si ricorda che la produzione è nulla quando X

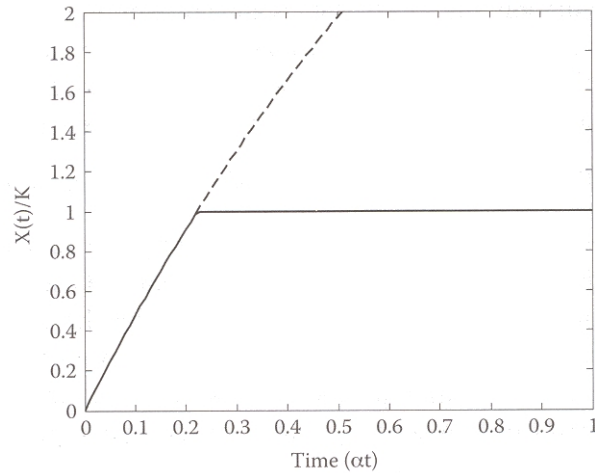


Figura 3.4: La dinamica della produzione di un gene autoregolato negativamente

supera K) (fig 3.4). Piccole oscillazioni si verificano attorno a $X = K$ se ci sono altri ritardi nel sistema. I ritardi possono causare l'overshoot di X superando leggermente la soglia di K . Subito dopo la produzione si ferma e il livello di X cala finché non diminuisce sotto quello di K , riprendendo così la produzione. In questo modo, X si stabilizza a un livello di steady state uguale al coefficiente di repressione del proprio promotore: $X_{st} = K$. L'andamento risultante mostra una salita ripida e una saturazione istantanea come in figura 3.4. Il tempo di risposta, $T_{1/2}$, può essere ricavato ponendo X uguale a metà dell'ampiezza dello steady state $X(T_{1/2}) = X_{st}/2$. Per semplicità, calcoliamo il tempo di risposta usando un'approssimazione lineare di X , con $X = \beta \cdot t$.

Il tempo di risposta $T_{1/2}^{n.a.r.}$, dove n.a.r. è l'autoregolazione negativa, è dato da X quando raggiunge metà del livello steady state:

$$\beta T_{1/2} = \frac{X_{st}}{2} = \frac{k}{2} \quad (3.11)$$

con

$$T_{1/2} = \frac{K}{2\beta} \quad \text{tempo di risposta per autoregolazione negativa} \quad (3.12)$$

Maggiore è la massima attività del promotore non represso, β , e minore è il tempo di risposta. L'autoregolazione negativa può usare un promotore forte per dare una produzione iniziale molto veloce, e poi usa la sua repressione per stabilizzare la produzione a un determinato livello. La soglia di repressione K può essere modificata, ad esempio, da mutazioni nel sito di legame di X nel

promotore, mentre β può essere modificato da mutazioni nel sito di legame dell'RNAP nel promotore. Così, la soglia $X_{st} = K$ e i tempi di risposta in principio possono essere determinati separatamente.

Confrontiamo questo modello con quello della regolazione semplice di un gene (senza autoregolazione negativa), il quale ha flusso di produzione $\beta_{semplice}$ e flusso di degradazione $\alpha_{semplice}$. Per avere un confronto significativo, dobbiamo confrontare i 2 modelli con lo stesso livello di soglia. In aggiunta, i modelli dovrebbero avere più parametri biochimici possibili simili tra di loro. Non a caso i due modelli confrontati hanno entrambe il flusso di utilizzazione $\alpha = \alpha_{semplice}$.

Nella regolazione semplice di un gene, lo steady state è il risultato del bilancio tra produzione e degradazione $X_{st} = \frac{\beta_{semplice}}{\alpha_{semplice}}$. In contrasto, come abbiamo appena visto, nell'autoregolazione negativa, lo steady state è dato dalla soglia di repressione $X_{st} = K$. Possiamo modificare K per avere entrambi i modelli con lo stesso livello di steady state: $K = \frac{\beta_{simple}}{\alpha_{simple}}$.

Il tempo di risposta della regolazione semplice è dato dalla relazione $T_{1/2} = \frac{\log 2}{\alpha_{simple}}$. Una risposta molto più veloce può essere ottenuta usando un'autoregolazione negativa e usando β molto grande, perché il tempo di risposta è $T_{1/2} = \frac{K}{2\beta}$, ed è inversamente proporzionale a β . Usando $K = \frac{\beta}{\alpha}$, troviamo che il rapporto fra i due tempi di risposta può essere molto piccolo incrementando β :

$$\frac{T_{1/2}^{n.a.r}}{T_{1/2}^{simple}} = \frac{\beta_{simple}/\beta}{2 \log 2} \quad (3.13)$$

La diminuzione dei tempi di risposta della rete di autoregolazione negativa rispetto alla rete senza autoregolazione è stata sperimentalmente dimostrata (dal gruppo di ricerca di Uri Alon) usando misurazioni ad alta risoluzione della dinamica dell'espressione dei geni (fig 3.5). In conclusione, l'autoregolazione negativa ha entrambi i vantaggi: un forte promotore dà una rapida produzione e il coefficiente di repressione dà il livello di steady state desiderato. Lo stesso promotore forte in una rete senza autoregolazione raggiunge un livello di steady state più elevato portando ad una sovrapproduzione del prodotto del gene.

3.4.1 L'autoregolazione negativa porta robustezza ai flussi di produzione

In aggiunta alla diminuzione dei tempi di risposta, l'autoregolazione negativa offre un secondo vantaggio: la diminuzione delle oscillazioni del flus-

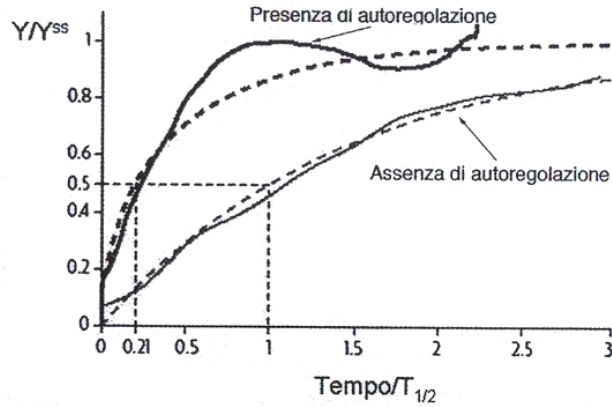


Figura 3.5: Misurazione sperimentale del motivo di regolazione in una rete reale

so di produzione β nello steady state.² Il flusso di produzione di un dato gene, β , oscilla nell'arco del tempo per via di tutte le oscillazioni presenti nel metabolismo della cellula e nei suoi sistemi regolatori e per via degli effetti stocastici della produzione di una proteina. Cellule gemelle solitamente hanno flussi di produzione diversi per la maggior parte delle proteine. La regolazione semplice di un gene è influenzata parecchio dalle variazioni del flusso di produzione β . Il livello dello steady state dipende linearmente dal suo flusso di produzione:

$$X_{st} = \frac{\beta}{\alpha} \quad (3.14)$$

quindi una variazione di β porta alla variazione in proporzione di X_{st} .

In contrasto, l'autoregolazione negativa può limitare le oscillazioni del flusso di produzione. Nel caso di un autorepressione ripida, il livello di steady state dipende solo dalla soglia di repressione di X per il proprio promotore $X_{st} = K$, determinata dai legami chimici tra X e il suo sito nel DNA. Questi parametri variano molto meno da cellula a cellula rispetto al flusso di produzione. Perciò l'autoregolazione negativa aumenta la robustezza dei livelli delle proteine nello steady state per quanto riguarda la maggior parte delle fluttuazioni, meglio ancora per la fluttuazione dei flussi di produzione.

²Questa proprietà è stata sperimentalmente dimostrata misurando il livello di proteine di singole cellule (Becskei and Serrano, 2000).

3.4.2 L'autoregolazione positiva

Alcuni fattori di trascrizione presentano autoregolazione positiva, in cui attivano la propria trascrizione. Questo si verifica, per esempio, in circa il 10% delle trascrizioni note dell'E. coli. L'autoregolazione positiva aumenta i tempi di risposta rispetto ad un sistema senza autoregolazione (fig 3.6). Inizialmente la dinamica è lenta, ma poi più il livello di produzione di X aumenta, più il suo flusso di produzione aumenta grazie alla retroazione positiva. Questo porta alla presenza di una curva concava finché non si raggiunge il 50% del valore del suo steady state relativo alla regolazione semplice. In questo modo l'autoregolazione positiva ha un effetto che è opposto a quello dell'autoregolazione negativa, ritardando i tempi di risposta. L'andamento lento dell'autoregolazione positiva può essere utile nei processi che richiedono tempi lunghi, come ad esempio il processo di sviluppo di un organismo.

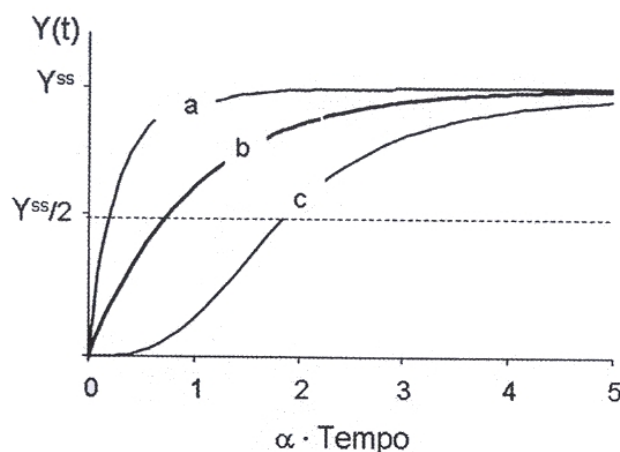


Figura 3.6: Andamento della concentrazione di una proteina: a) autoregolazione negativa, b) assenza di autoregolazione, c) autoregolazione positiva

In aggiunta, quando il flusso dell'autoregolazione positiva è più elevato rispetto a quello dell'utilizzazione, il sistema può diventare bistabile. Una volta che il gene è stato attivato, resta bloccato in uno stato di elevata espressione e si mantiene sempre attivo, anche dopo che l'attivazione iniziale è svanita (Carrier and Keasling, 1999; Demongeot et al., 2000; Becskei et al., 2001; Ferrell, 2002; and Isaacs et al., 2003). Questo tipo di circuito di memoria viene usato nelle reti di trascrizione dello sviluppo per prendere

decisioni irreversibili come ad esempio la determinazione del tipo di tessuto che la cellula avrà in un organismo multicellulare.

Capitolo 4

Regolazione feed forward

4.1 Introduzione

In questo capitolo vengono analizzati i moduli di regolazione con tre nodi, osservando se possono essere dei motivi o meno, discutendo anche la loro funzionalità. Per trovare pattern importanti, prima si calcola il numero di possibile di combinazioni e con considerazioni probabilistiche analoghe a quelle del capitolo precedente si calcola la loro frequenza in una rete random e poi viene confrontata con una rete reale. Una volta trovato il motivo più ricorrente, cioè la regolazione feed forward, viene analizzata la sua dinamica per capire quali funzioni utili ha.

4.2 Motivi con 3 nodi

Analizzando i moduli di regolazione con 3 nodi, si può osservare che ci sono 13 combinazioni possibili come in figura 4.1. Nella seguente tabella sono stati calcolati i valori della regolazione feed-forward (fig. 4.2a) e della regolazione feedback (fig. 4.2b) sia nella rete random che in quella reale dell'E. coli[1]. Si può osservare chiaramente che la rete della regolazione feed-forward è chiaramente un motivo, essendo presente 42 volte, mentre la regolazione feedback è assente.

	Motivi feed forward	Motivi feedback
E. coli	42	0
Reti random ER	1.7 ± 1.3	0.6 ± 0.8

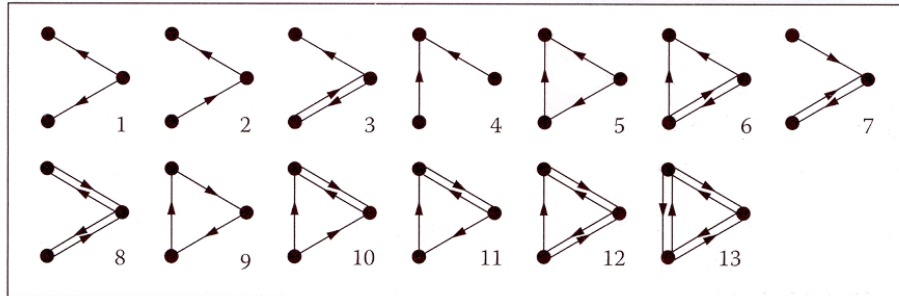


Figura 4.1: Possibili combinazioni dei moduli di regolazione con 3 nodi

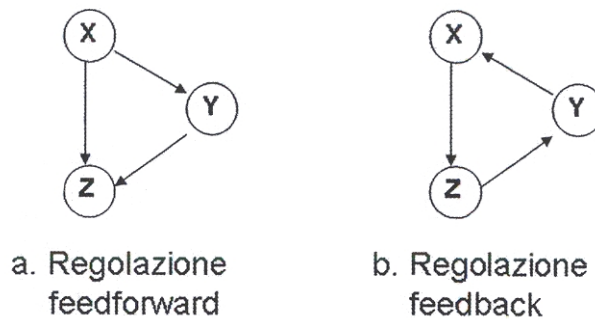


Figura 4.2: a) Regolazione feed forward, b) regolazione feedback

4.3 Feed forward

Il motivo di regolazione presenta un fattore di trascrizione X che regola un altro fattore Y . Assieme X e Y regolano la produzione della proteina Z . Quindi la regolazione tra X e Z avviene con due meccanismi, uno diretto e uno tramite il fattore Y . Poiché ai 3 archi $X \rightarrow Y$, $Y \rightarrow Z$, $X \rightarrow Z$ può corrispondere un'attivazione o una repressione, abbiamo $2^3 = 8$ possibili combinazioni nella stessa rete. Le varie combinazioni possono essere divise in 2 categorie: *coerenti* quando le 2 vie, diretta ($X \rightarrow Z$) e indiretta ($X \rightarrow Y$, $Y \rightarrow Z$), hanno lo stesso effetto, o di attivazione o di repressione, e *incoerenti* nel caso contrario, cioè nel caso in cui le 2 vie hanno effetti opposti. Tutte le 8 possibili combinazioni non compaiono con la stessa frequenza, in particolare

la più frequente è di tipo coerente e ha tutti i meccanismi di attivazione presenti in tutti i 3 gli archi.

Per simulare il comportamento di questo motivo, supponiamo che sia necessaria la presenza contemporanea dei due fattori X e Y per produrre Z . Questo può essere descritto dalla funzione logica AND tra X e Y . I due fattori rispondano a due stimoli esterni S_x e S_y , passando dalla forma inattiva alla forma attiva X^* e Y^* . Lo schema è rappresentato nella fig. 4.3.

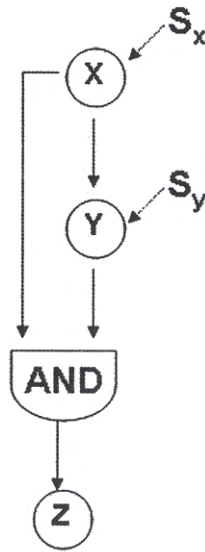


Figura 4.3: Motivo di regolazione feed forward

Il modello matematico del circuito è il seguente:

$$\frac{dY(t)}{dt} = \frac{\beta_1 X^*(t)^{n_1}}{K_1 + X^*(t)^{n_1}} - \alpha_1 Y(t) \quad (4.1)$$

$$\frac{dZ(t)}{dt} = \frac{\beta_2 X^*(t)^{n_2}}{K_2 + X^*(t)^{n_2}} \frac{Y^*(t)^{n_3}}{K_3 + Y^*(t)^{n_3}} - \alpha_2 Z(t) \quad (4.2)$$

Per un'analisi semplificata in modo da comprendere intuitivamente il funzionamento del circuito, approssimiamo le funzioni di Hill con funzioni gradino:

$$\frac{\beta_1 X^*(t)^{n_1}}{K_1 + X^*(t)^{n_1}} = \begin{cases} \beta_1 & \text{se } X^*(t) \geq K_1 \\ 0 & \text{se } X^*(t) < K_1 \end{cases} \quad (4.3)$$

$$\frac{\beta_2 X^*(t)^{n_2}}{K_2 + X^*(t)^{n_2}} \frac{Y^*(t)^{n_3}}{K_3 + Y^*(t)^{n_3}} = \begin{cases} \beta_2 & \text{se } X^*(t) \geq K_2 \text{ AND } Y^*(t) \geq K_3 \\ 0 & \text{altrimenti} \end{cases} \quad (4.4)$$

I valori soglia sono definiti dai parametri K_1 , K_2 e K_3 . Per esempio, la produzione della proteina Y avviene alla velocità β solo se la concentrazione del fattore di produzione nella forma attiva X^* supera la soglia K_1 . Supponendo che all'istante $t = 0$, in seguito alla risposta al segnale S_x , X^* superi la soglia K_1 come in figura 4.4. Per effetto della produzione β , Y sale esponenzialmente verso il valore di soglia $\frac{\beta_1}{\alpha_1}$:

$$Y(t) = \frac{\beta_1}{\alpha_1} (1 - e^{-\alpha_1 t}) \quad (4.5)$$

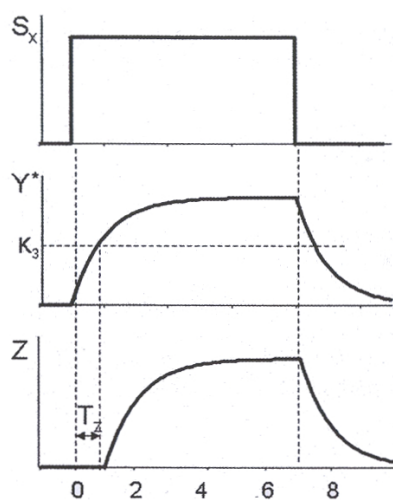


Figura 4.4: Andamento temporale delle concentrazioni di Y e Z

Poiché la produzione di Z richiede la presenza contemporanea di dei due fattori X e Y nella loro forma attiva, con la concentrazione di X^* maggiore di K_2 e quella di Y maggiore di K_3 , è necessario che sia sufficientemente intenso anche il segnale S_y oltre a S_x . La produzione di Z inizia quando Y^* supera la soglia K_3 la tempo T_z . Per calcolare T_z basta porre l'espressione analitica di Y uguale a K_3 :

$$Y^*(T_z) = \frac{\beta_1}{\alpha_1} (1 - e^{-\alpha_1 T_z}) = K_3 \quad (4.6)$$

Come ci si poteva aspettare T_z dipende dai parametri biochimici della proteina Y ed è tanto maggiore quanto più α_1 è basso e quanto più K_3 tende al valore di soglia $\frac{\beta_1}{\alpha_1}$:

$$T_z = \frac{1}{\alpha_1} \log \frac{\frac{\beta_1}{\alpha_1}}{\frac{\beta_1}{\alpha_1} - K_3} \quad (4.7)$$

Dall'istante T_z in poi la concentrazione di Z sale con un andamento esponenziale:

$$Z(t) = \frac{\beta_2}{\alpha_2} (1 - e^{-\alpha_2(t-T_z)}) \quad (4.8)$$

La proteina Z non viene mai prodotta se la soglia di attivazione K_3 è superiore al valore di regime $\frac{\beta_1}{\alpha_1}$. Nei circuiti reali il valore di K_3 è nettamente inferiore a $\frac{\beta_1}{\alpha_1}$, permettendo alla produzione di Z di non risentire le possibili fluttuazioni nel tempo della concentrazione di Y^* , che continua a variare nel tempo sopra la soglia di K_3 .

Alla scomparsa del segnale S_x , X diventa rapidamente inattivo, portando istantaneamente all'arresto della produzione di Y e Z . Quindi, mentre c'è un ritardo di risposta di Z alla comparsa di S_x , non c'è ritardo quando S_x cessa.

Questo circuito permette di filtrare anche piccole variazioni nel segnale di ingressi: se il segnale S_x si presenta per un tempo inferiore a T_z , Y^* non raggiunge la soglia necessaria per superare K_3 . Soltanto degli stimoli durevoli provocano la comparsa di Z e questa proprietà è molto importante perché la produzione di Z non è influenzata da stimoli di breve durata, come ad esempio per effetto di fluttuazioni random (figura 4.5).

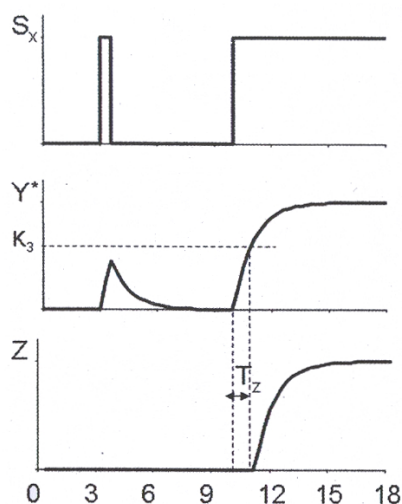


Figura 4.5: Risposta a stimoli del motivo di regolazione feed forward

Il motivo di regolazione appena analizzato è presente nel sistema arabinoso dell'E. coli (Mangan et al., 2003), cioè nel sistema che produce enzimi in grado di trasportare dentro la cellula un tipo di zucchero, arabinoso, che poi viene utilizzato per produrre energia.

Conclusioni

Come si è potuto vedere, nei sistemi biologici l'espressione genica non è un indicatore sufficiente dell'espressione proteica. Infatti l'analisi del trascrittoma ci dà solo un'indicazione di quali geni siano attivi, ma nulla sull'espressione proteica, la quale oltre a dipendere da un numero elevato di geni, trascritti, proteine e metaboliti, dipende anche dalle alterazioni indotte dal sistema.

Per poter descrivere il funzionamento complesso di una cellula si usano reti di trascrizione molto articolate, che presentano un elevato numero di interazioni tra fattori di trascrizione e geni rendendo il loro studio difficile. Per semplificare l'analisi vengono usati modelli semplici isolati dal sistema. Si inizia con l'analisi del turnover delle singole proteine per poi approfondire la regolazione della trascrizione dei geni.

Per analizzare il turnover delle singole proteine bisogna integrare le informazioni sui livelli di espressione genica e proteica con informazioni sul loro comportamento dinamico, evidenziando come la concentrazione di una proteina nel plasma è funzione dalla sua velocità di produzione e dalla sua velocità di degradazione. Questa analisi viene ulteriormente semplificata dalla diversa scala dei tempi per i vari fenomeni che si verificano all'interno delle reti di trascrizione, permettendo di assumere i livelli di attività dei fattori di trascrizione come costanti nelle equazioni che descrivono la dinamica della rete sulla scala dei tempi del cambiamento dei livelli proteici.

L'analisi del sistema durante una perturbazione ha evidenziato come le reti biologiche possono essere studiate con un certo grado di semplicità in quanto presentano un numero molto minore di pattern rispetto a tutti quelli possibili presenti in una rete random. Un numero così piccolo è dovuto alla selezione naturale che impone alle reti di trascrizione di rispettare dei requisiti di base quali la robustezza rispetto a fluttuazioni random dei segnali

di ingresso e l'efficienza, svolgendo funzioni specifiche con il minimo numero di componenti.

Un'analisi teorica delle proprietà di alcuni pattern ricorrenti come i motivi di autoregolazione negativa e positiva o il feed-forward, permette di descrivere le loro caratteristiche all'interno di una rete reale, come tempi di risposta molto brevi per l'autoregolazione negativa o molto lunghi per l'autoregolazione positiva e immunità a stimoli di breve durata per il feed-forward.

Con studi futuri sarà interessante approfondire il comportamento dei motivi non isolati, ma collegati tra di loro per vedere come essi influiscono complessivamente sul funzionamento di una rete reale e se un'analisi approfondita di motivi isolati porta a capire bene il funzionamento della rete senza doverla affrontare nel suo complesso.

Bibliografia

- [1] Alon U. An introduction to System Biology CRC Press Taylor and Francis Group, 2006
- [2] Barbarini N. e Magni P. Analisi di dati di spettrometria di massa finalizzata alla scoperta di biomarcatori, Genomica e proteomica computazionale Pàtron editore, 2007
- [3] Erdos, P. e A. Renyi, On random graphs, pubblicazione matematica Debrecen, 1959
- [4] Shen.Orr, S.S., R. Milo, S. Mangan e U. Alon, Network motifs in the transcriptional regulation network of Escherichia coli, Nature Genet 31: 64-68, 2002
- [5] G. Toffolo, B. Di Camillo e C. Cobelli Modelli del turnover e della regolazione proteica, Genomica e proteomica computazionale Pàtron editore, 2007