

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

---

Scuola di Scienze  
Corso di Laurea in Fisica

# STUDIO DEL NETWORK METABOLICO UMANO: MODELLI ED APPLICAZIONI

Relatore:  
Prof. Daniel Remondini

Presentata da:  
Gabriele Fontanarosa

Sessione III  
Anno Accademico 2014/2015

## Abstract

Il lavoro di tesi si basa sull'analisi del database pubblico ReconX, il quale costituisce lo stato dell'arte della ricostruzione in silico del metabolismo umano. Il modello che se ne può estrarre è stato impiegato per simulazioni di metabolismo cellulare in condizioni di ossigeno variabile, valutando l'impatto della carenza di ossigeno sulle reazioni e sui pathways metabolici. Le tecniche impiegate appartengono alla systems biology e sono di tipo bioinformatico e riguardano flux balance analysis e flux variability analysis. I risultati trovati vengono quindi confrontati con la letteratura di settore. È stato inoltre possibile estrarre dei sotto network dal modello principale, analizzando la rete di connessioni esistente fra i metaboliti e fra le reazioni separatamente. Viene estratto e studiato anche il network di interazione fra pathways, su cui è introdotta una misura ad hoc per valutare la forza di connessione fra i vari processi. Su quest'ultimo network viene anche effettuata un'analisi di tipo stocastico, mostrando che tecniche di tipo markoviano possono essere applicate per indagini diverse da quelle più in uso basate sui flussi. Infine viene mostrata una possibile soluzione per visualizzare graficamente i network metabolici così costituiti.



*“How silly is that? A heart cell hating a lung cell.”*

*Cassie from **THE THREE***



# Introduzione

L'applicazione sistematica della fisica alla fisiologia è tutto sommato storia recente: ad esclusione delle prime osservazioni ed applicazioni di metodi ancora molto rudimentali (le sperimentazioni elettriche di Galvani nel '700, la presa di coscienza di flussi sanguinei nell'organismo del '600) bisogna aspettare la fine del '800 perché i risultati della ricerca fisica diventino uno strumento di analisi della materia vivente. I raggi x e la radioattività sono attivamente utilizzati tutt'oggi a scopo diagnostico (tac, pet, nmr...) che terapeutico (radioterapia con acceleratori lineari) e di grande importanza sono le applicazioni laser per interventi di varia natura. Queste tecniche costituiscono tuttavia la stragrande maggioranza dell'impatto della fisica più recente sulla fisiologia e medicina. Negli ultimi anni si assiste ad un nuovo approccio delle scienze dure alla biologia, stavolta indirizzato non alla realizzazione di strumenti di diagnosi e analisi ma alla comprensione e simulazione dei fenomeni molecolari che avvengono all'interno delle cellule: tale sforzo ha lo scopo di consentire alla fisica, alla matematica e all'informatica di costruire e validare teorie con capacità modellistiche e predittive, entrando così a pieno titolo fra gli strumenti in grado di progredire la scienza della materia vivente. L'insieme delle tecniche per la modellizzazione di strutture biologiche complesse prende il nome di biologia dei sistemi, branca nella quale confluiscono teorie e metodi atti ad uno studio meno riduzionistico di ciò che avviene a livello molecolare: i sempre più abbondanti dati omici (genomica, proteomica e metabolomica), la potenza di calcolo offerta dai computer e il background fisico-matematico nello studio dei sistemi complessi concorrono insieme a de-

lineare l'approccio olistico della biologia dei sistemi sul vivente. A partire dalla codifica dei geni, proseguendo con la catalogazione degli enzimi che ne dipendono fino ad arrivare agli interi networks di funzioni cellulari, lo scopo della biologia dei sistemi è quello di fornire l'insieme dei principi teorici che regolano il mondo del vivente.

Questo lavoro di tesi si inserisce nell'ambito di sovrapposizione della biologia dei sistemi fra metabolismo cellulare, la fisica delle reti, la matematica e la bioinformatica. La prima parte di questo lavoro riguarda l'introduzione delle ricostruzioni metaboliche, che costituiscono la base descrittiva su cui appoggeremo la nostra analisi. Successivamente verranno introdotti i concetti matematici e di teoria delle reti necessari alla manipolazione di queste ricostruzioni per discutere infine un'applicazione pratica di flux balance analysis e modelling markoviano sullo stato dell'arte della ricostruzione del metabolismo umano.

# Indice

<b>Introduzione</b>	<b>i</b>
<b>1 Le ricostruzioni metaboliche</b>	<b>1</b>
1.1 Cenni sul metabolismo . . . . .	2
1.2 I pathways metabolici . . . . .	4
1.3 I modelli in silico del metabolismo . . . . .	6
1.4 La matrice stechiometrica . . . . .	9
<b>2 Flux Balance Analysis e Flux Variability</b>	<b>13</b>
2.1 Ipotesi sul sistema . . . . .	14
2.2 Programmazione Lineare . . . . .	19
2.3 Applicazioni e limiti del FBA: la Flux Variability Analysis . . . . .	20
<b>3 Recon 2</b>	<b>25</b>
3.1 Le caratteristiche della ricostruzione . . . . .	27
3.2 Reazione di biomassa umana . . . . .	29
3.3 La rete dei metaboliti e delle reazioni . . . . .	31
3.4 Misure di centralità . . . . .	32
<b>4 Il network dei pathways</b>	<b>39</b>
4.1 Il network pathway-pathway . . . . .	40
4.2 Misura alternativa per la forza delle connessioni . . . . .	41
4.3 Analisi stocastica . . . . .	43

---

<b>5</b>	<b>Un'applicazione FBA-FVA: ipossia cellulare</b>	<b>47</b>
5.1	Metabolismo delle cellule cancerose: effetto Warburg . . . . .	48
5.2	L'ipossia nel modello Recon2 . . . . .	51
5.3	Cobra toolbox e FBA . . . . .	52
5.4	Introduzione della FVA nell'indagine . . . . .	53
5.5	Simulazione e risultati . . . . .	55
	<b>Conclusioni</b>	<b>63</b>
	<b>A Tabelle di supporto</b>	<b>67</b>
	<b>Bibliografia</b>	<b>73</b>

# Elenco delle figure

1.1	Funzionamento di un enzima . . . . .	3
1.2	Visione d'insieme del metabolismo . . . . .	4
1.3	Schema del tca cycle . . . . .	5
1.4	Costruzione di un modello metabolico . . . . .	7
1.5	Risorse online . . . . .	8
1.6	Dalle reazioni alla matrice stechiometrica . . . . .	10
1.7	Sottospazi fondamentali della matrice S . . . . .	11
2.1	Schema FBA . . . . .	17
2.2	Ottimizzazione MOMA . . . . .	18
2.3	Dal network metabolico alla soluzione FBA . . . . .	21
2.4	Visualizzazione risultato FBA . . . . .	22
3.1	Mappa del metabolismo umano . . . . .	25
3.2	Lo sviluppo di Recon2 . . . . .	26
3.3	Criterio per la creazione dei link nella rete dei metaboliti . . . . .	32
3.4	Criterio per la creazione dei link nella rete delle reazioni . . . . .	33
3.5	Distribuzione connettività network metabolita-metabolita . . . . .	34
3.6	Distribuzione della connettività nel network reazione-reazione . . . . .	35
4.1	Criterio per la creazione dei link nella rete dei pathway . . . . .	40
4.2	Distribuzione dei link pesati nella rete pathway-pathway e valore del link con più peso per pathway . . . . .	41
4.3	Distribuzione connettività network reazione-reazione . . . . .	45

---

5.1	Respirazione cellulare . . . . .	49
5.2	Vasi sanguinei, stroma e tessuto tumorale . . . . .	50
5.3	Numero di reazioni con flusso diminuito e aumentato per run .	56
5.4	Pathways con reazioni a flusso aumentato per run . . . . .	58
5.5	Numero pathways con reazioni a flusso diminuito lungo i run .	60
5.6	Rapporto percentuale fra il flusso totale delle reazioni a flusso diminuito e baseline . . . . .	60
5.7	Percentuale di reazioni a flusso calato rispetto il numero totale di reazioni per pathways, lungo tutti i run . . . . .	61

# Elenco delle tabelle

3.1	Alcuni dei campi presenti nella struttura matlab di Recon2 . . .	27
3.2	Composizione della reazione di biomassa umana: metaboliti con i loro coefficienti stechiometrici. In grassetto i 3 metaboliti prodotti dalla reazione. . . . .	30
3.3	I dieci metaboliti a connettività più alta . . . . .	34
3.4	Le dieci reazioni a connettività più alta . . . . .	35
3.5	I dieci metaboliti a betweenness centrality più alta, normalizzata rispetto il valore della prima . . . . .	37
3.6	Le dieci reazioni a betweenness centrality più alta, normalizzata rispetto il valore della prima . . . . .	37
4.1	Top ranking pathways secondo la misura della forza di connessione. . . . .	42
4.2	Top ranking pathways secondo il vettore dello stato stazionario. . . . .	44
5.1	Differenze valori flussi fra glpk e gurobi . . . . .	53
5.2	Convergenza dei risultati FVA alla variazione della tolleranza . . . . .	55
5.3	Paths con reazioni a flusso aumentato in almeno 1000 run . . . . .	57
5.4	Pathways con reazioni a flusso diminuito in tutti i 2000 run . . . . .	59
A.1	Pathways di recon, con numero reazioni e metaboliti unici. . . . .	67
A.2	Reazioni di Recon2 interessate dall'ipossia . . . . .	71
A.3	Reazioni di Recon2 non interessate dall'ipossia . . . . .	72



# Capitolo 1

## Le ricostruzioni metaboliche

La biologia della seconda metà del XX secolo è stata influenzata dall'approccio riduzionistico, concentrandosi sullo studio e la generazione di informazioni riguardanti componenti cellulari individualmente, sulla loro composizione chimica e sulla loro funzione fisiologica. A partire da metà degli anni '90 questo processo è stato accelerato dalla nascita della genomica. Allo stato attuale è stato sequenziato il DNA (l'acido nucleico che contiene le informazioni genetiche necessarie alla biosintesi di RNA e proteine alla base dei sistemi viventi) di un numero sempre crescente di organismi, di cui si sta mano a mano completando l'intero patrimonio. Nonostante per ora l'assegnazione di una o più funzioni per i geni codificanti di questi organismi non sia completa, ci si aspetta che questo processo possa concludersi per la maggior parte di essi in futuro. Inoltre, l'incrocio di dati a partire da differenti genomi accelererà questo sviluppo, portando alla compilazione più precisa di un catalogo di componenti cellulari per un grande numero di organismi. Come possiamo utilizzare questa crescente lista di oggetti cellulari e rispettive proprietà? Per quanto informative questo genere di raccolte forniscono unicamente le informazioni basilari riguardo le molecole che compongono le cellule, le loro specifiche proprietà chimiche e quando queste scelgono di usare i loro componenti. Come possiamo arrivare alle proprietà biologiche e di comportamenti che esulano da queste liste dettagliate? Queste proprietà

emergenti si presentano quando si considera la totalità dei processi che caratterizzano la cellula. La biologia dei sistemi si concentra sulla natura dei collegamenti generati fra le varie parti di una cellula ed i loro costituenti, studiando la rete globale che si viene a creare in tal modo.

## 1.1 Cenni sul metabolismo

Con metabolismo si intende il set di trasformazioni chimiche atte al sostenimento della vita cellulare di un organismo: le reazioni che compongono questo set consentono all'organismo di crescere, riprodursi, mantenere le proprie strutture e rispondere agli stimoli ambientali. Il metabolismo viene generalmente suddiviso in due insieme di processi:

- catabolismo, responsabile della degradazione delle particelle complesse e produzione dell'energia sotto forma di molecole di ATP tramite respirazione cellulare, le reazioni di questo tipo sono spontanee
- anabolismo, che a partire da componenti più semplici costruisce molecole complesse come proteine (fra cui gli enzimi), acidi nucleici etc. Le reazioni di tipo anabolico non sono spontanee e richiedono l'intervento di enzimi specifici.

Le reazioni chimiche del metabolismo sono organizzate nei cosiddetti pathways metabolici, in cui un componente chimico (metabolita) è trasformato lungo una serie di steps in un altro componente tramite l'intervento di enzimi. Gli enzimi sono per lo più proteine codificate dal DNA (ne esistono una minoranza composte da molecole di RNA dette ribozomi) e rappresentano un elemento cruciale al metabolismo in quanto permettono agli organismi di condurre reazioni che richiederebbero energie di attivazione superiori a quelle disponibili spontaneamente (fig 1.1). Il substrato costituito dalle molecole della reazione da svolgere si inserisce nel sito attivo dell'enzima: l'energia di attivazione si abbassa e la reazione può compiersi. L'enzima non prende parte al processo chimicamente, ne si consuma al suo completamento. Gli

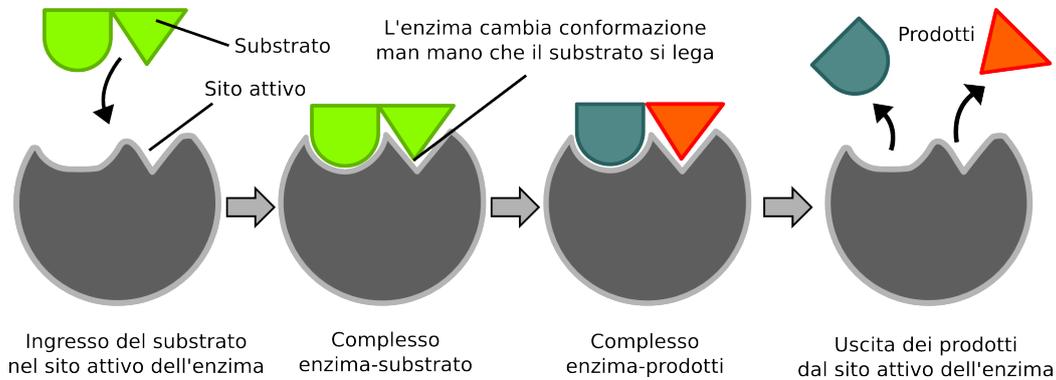


Figura 1.1: Schema di funzionamento di un enzima: il substrato, la reazione da catalizzare, si inserisce nel sito attivo dell'enzima. La reazione di separazione fra i due componenti può quindi avvenire.

enzimi agiscono come catalizzatori, accelerando il completamento della reazione e sono i primi responsabili della regolazione dei pathways metabolici in risposta ai cambiamenti dell'ambiente cellulare o a segnali da altre cellule. Un'importante caratteristica del metabolismo riguarda la similarità fra i pathways metabolici di base anche fra differenti specie: ad esempio l'insieme degli intermediari del ciclo dell'acido citrico sono presenti in tutti gli organismi conosciuti, trovandosi in specie diverse come i batteri unicellulari di *Escherichia Coli* e grossi organismi pluricellulari come gli elefanti [1]. Queste somiglianze sono molto probabilmente dovute alla loro comparsa per motivi evolutivisti e il loro mantenimento dovuto alla loro efficacia. Dal punto di vista termodinamico, gli organismi sono soggetti alla seconda legge (l'aumento di entropia) esattamente come qualsiasi altro sistema fisico: il metabolismo mette in comunicazione i sistemi viventi con il loro ambiente e permette lo scambio di energia e composti. Così facendo i sistemi biologici sono lontani dall'equilibrio e fisicamente costituiscono sistemi dissipativi in grado di mantenere la loro complessità ed ordine aumentando l'entropia dell'ambiente circostante. In termini termodinamici, il metabolismo mantiene ordine creando disordine a scapito dell'ambiente esterno.

## 1.2 I pathways metabolici

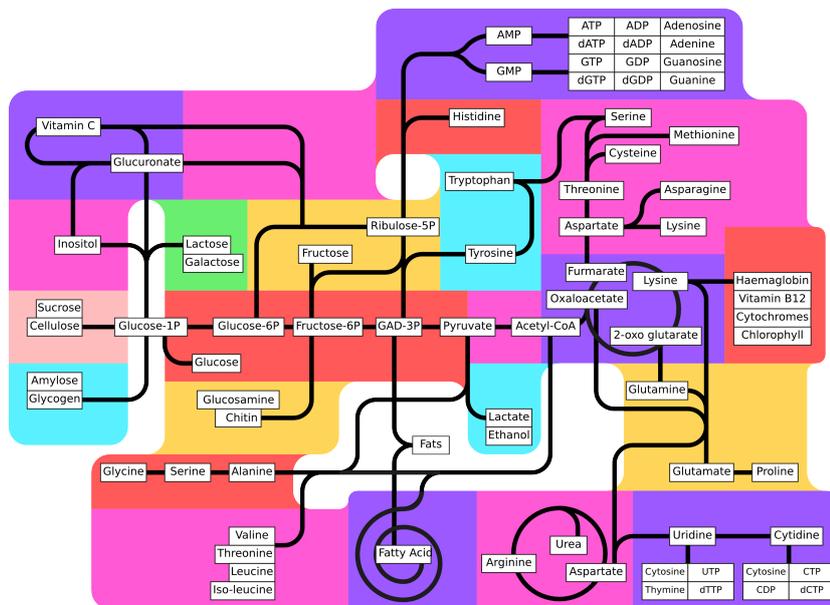


Figura 1.2: Visione d'insieme del metabolismo composto da pathways

Le reazioni che costituiscono il metabolismo devono essere costantemente regolate in base ai mutamenti dell'ambiente esterno: tali regolazioni costituiscono l'omeostasi, ossia l'insieme dei processi che mantengono la cellula in equilibrio. Ogni pathway consiste in una serie di reazioni biochimiche collegate dai loro prodotti intermedi in cui il risultato di una reazione è il substrato della successiva. La direzione di scorrimento di questa cascata di reazioni non è necessariamente univoco: in molti casi la termodinamica della singola reazione impone un senso di svolgimento e l'eventuale scomposizione del prodotto di una reazione è l'oggetto di uno step di un altro pathway. Altre volte può succedere che una reazione possa svolgersi in entrambi i sensi. Per esempio il primo pathway che è stato individuato, quello della glicolisi, ha come obiettivo la scomposizione del glucosio, ma molte reazioni che ne fanno parte vengono anche percorse al contrario portando a gluconogenesi.

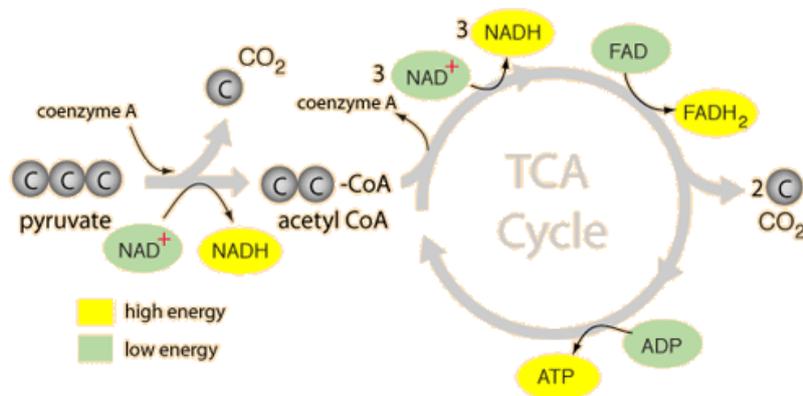


Figura 1.3: Esempio di pathway circolare: il tca cycle, o ciclo di Krebs, avviene negli eucarioti all'interno dei mitocondri ed ha come obiettivo la produzione di molecole energetiche di ATP (fonte: <http://hyperphysics.phy-astr.gsu.edu/>).

- I pathways metabolici sono regolati da feedback negativo, quindi la concentrazione dei prodotti delle reazioni e dei componenti presenti nella cellula influiscono sullo svolgimento successivo delle stesse.
- Alcuni pathway scorrono in circolo, in maniera che l'ultimo prodotto della catena sia substrato per la prima reazione che la compone. Un esempio potrebbe essere il ciclo di Krebs, figura 1.3.
- Le reazioni cataboliche ed anaboliche spesso avvengono separatamente ed in maniera indipendente, confinate fisicamente in organelli diversi o chimicamente perché catalizzate da enzimi diversi.

Molti pathways possono coesistere all'interno di una stessa cellula, creando un network nel momento in cui i prodotti di uno diventano i reagenti di un altro.

## 1.3 I modelli in silico del metabolismo

Come abbiamo brevemente visto, il metabolismo è un sistema strutturato a più livelli che possiamo ordinare dal più generale al più particolare:

**input/output cellulare** : il metabolismo è visto come una complessa macchina che trasforma nutrienti in biomolecole per la crescita cellulare e i rifiuti sono espulsi dalla cellula nell'ambiente

**catabolismo e anabolismo** : i due settori principali del metabolismo, uno destinato al processamento dei nutrienti e l'altro per la biosintesi di macromolecole.

**pathways** : ad una risoluzione più precisa, i pathways sono identificati come un set di reazioni che servono a uno scopo definito (sintesi dei lipidi, degradazione carboidrati complessi etc.)

**le singole reazioni** : ogni trasformazione può formare o rompere un legame chimico, può trasferire un gruppo chimico da un gruppo di molecole ad un altro, può ossidare o ridurre un composto chimico, produrre o consumare energia etc. Tranne che per reazioni molto semplici, tutti questi processi sono catalizzati da un enzima la cui presenza o meno determina lo svolgersi o meno della reazione. In accordo con il tipo di trasformazione catalizzata, gli enzimi e le conseguenti reazioni sono etichettate con il codice universale EC (Enzyme Commission).

Dalla loro introduzione a metà degli anni novanta, il compito delle ricostruzioni metaboliche è smontare i pathways nei loro rispettivi enzimi e reazioni raccogliendo tutte le informazioni metaboliche rilevanti e compilandole in un modello matematicamente coerente. La potenza di questi modelli risiede nella loro capacità di corrispondere ad un gene il rispettivo pathway metabolico e la loro qualità dipende dalla capacità di inferire un fenotipo a partire dal genotipo. I dati necessari alla costruzione di questi modelli arrivano da diversi settori:

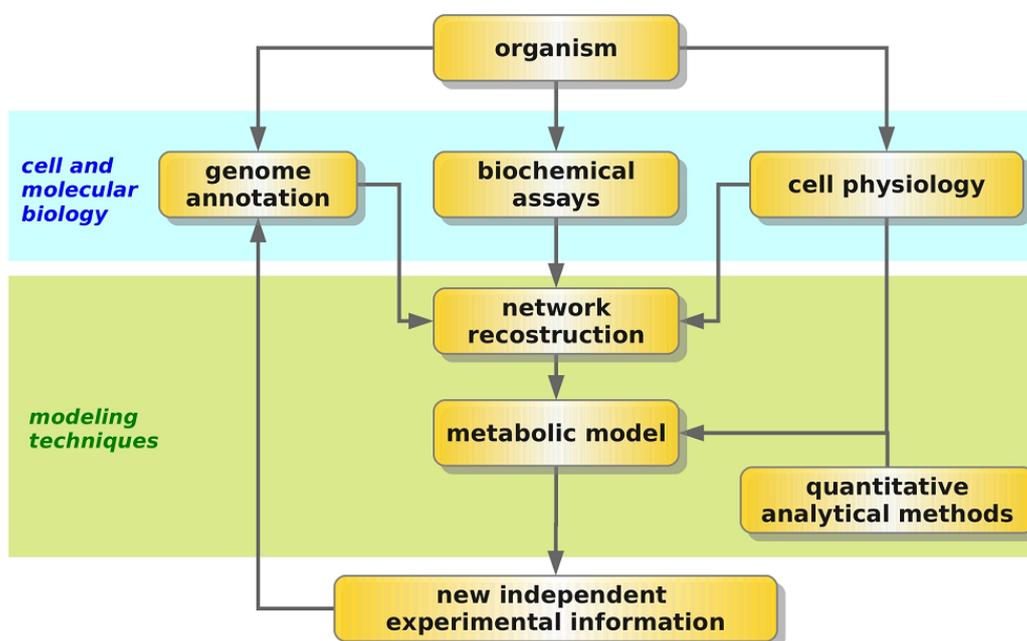


Figura 1.4: Schema di acquisizione dati per la costruzione di un modello metabolico. (Fonte: [2])

- biochimica: l'evidenza più forte dell'esistenza di una reazione metabolica deriva dalla possibilità di isolarne il corrispondente enzima;
- genomica: sulla base della sequenza di DNA è possibile procedere ad assegnazioni funzionali per i geni e relative reazioni;
- fisiologia: ad esempio, l'abilità per un organismo di produrre un amminoacido potrebbe supportare l'ipotesi dell'esistenza della rispettiva via metabolica (pathway);
- modeling in silico: per identificare un gap nella catena di reazioni che portano ad un componente, un studio in silico può essere utile nel tentativo di inferire posizione e ruolo dello step mancante.

Sinteticamente, il processo per compilare una ricostruzione metabolica è il seguente (figura 1.4):

1. prima ricostruzione draft, a partire da database (KEGG, BRENDA, Entrez Gene, PubChem...) e pubblicazioni: in questa fase vengono raccolte tutte le informazioni disponibili online riguardo le vie e i geni metabolici dell'organismo di cui si vuole costruire un modello (figura 1.5);
2. raffinamento del modello: la prima ricostruzione viene aggiornata ed integrata da dati di laboratorio quando disponibili;
3. conversione del modello in una rappresentazione matematica e computabile: a questo punto è possibile creare la struttura matematica che consentirà la manipolazione della ricostruzione;
4. valutazione e debug del modello tramite sperimentazione per valutarne l'accordo: in questa fase il modello viene messo alla prova con sperimentazioni ed osservazioni di laboratorio per testarne l'efficacia. Sarà sempre possibile aggiungere e correggere la ricostruzione per aggiornarla a nuovi risultati.

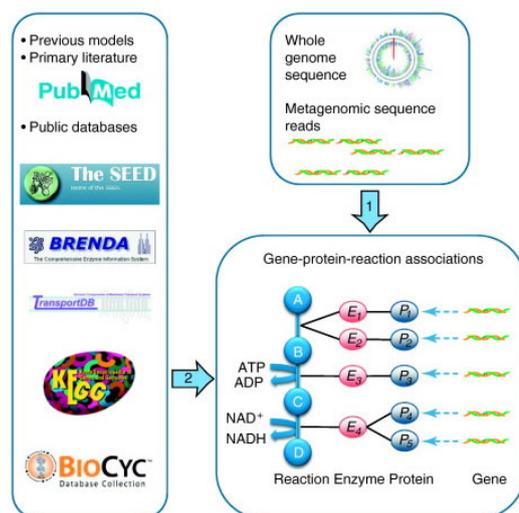


Figura 1.5: Integrazione delle risorse online per la compilazione del modello metabolico.

## 1.4 La matrice stechiometrica

Un set di reazioni chimiche può essere schematizzato con un insieme di equazioni chimiche, al cui interno sono contenute le informazioni riguardanti la stechiometria delle reazioni stesse. Queste informazioni stechiometriche possono essere rappresentate a loro volta in forma matriciale, matrice che prende il nome di matrice stechiometrica, che verrà indicata con  $S$ . Associata a questa matrice si possono trovare informazioni sugli enzimi, livelli di trascrizione, annotazioni etc...Una volta compilata, la matrice stechiometrica rappresenta un database biochimico, genetico e genomico strutturato ed è infatti il cuore di ogni ricostruzione metabolica così come sono state presentate. Nonostante  $S$  sia il risultato di una raccolta dati di tipo biochimico, essa è anche un oggetto di tipo matematico e può essere analizzata per determinare le proprietà del network di reazioni che descrive. La matrice è formata dai coefficienti interi stechiometrici delle reazioni ed è organizzata in maniera tale che ad ogni colonna corrisponda ad una reazione e ogni riga corrisponda ad un composto. Quindi ogni riga indica le reazioni in cui un particolare metabolita prende parte, mostrando come fra di loro le reazioni sono connesse. Per convenzione, i metaboliti reagenti di una reazione avranno un coefficiente di valore intero negativo, mentre i metaboliti prodotti da una reazione avranno un coefficiente di valore intero positivo.

Ogni matrice stechiometrica ha dimensioni  $M \times R$  dove  $M$  rappresenta il numero di metaboliti ed  $R$  il numero di reazioni e i coefficienti assumono valore positivo se sono i prodotti nella reazione e valore negativo se sono i reagenti. Tipicamente, per un network di reazioni biologico, troviamo più reazioni che metaboliti. Matematicamente, la matrice stechiometrica  $S$  è una trasformazione lineare del vettore dei flussi

$$v = (v_1, v_2 \dots v_n)$$

al vettore delle derivate temporali delle concentrazioni

$$x = (x_1, x_2 \dots x_n)$$

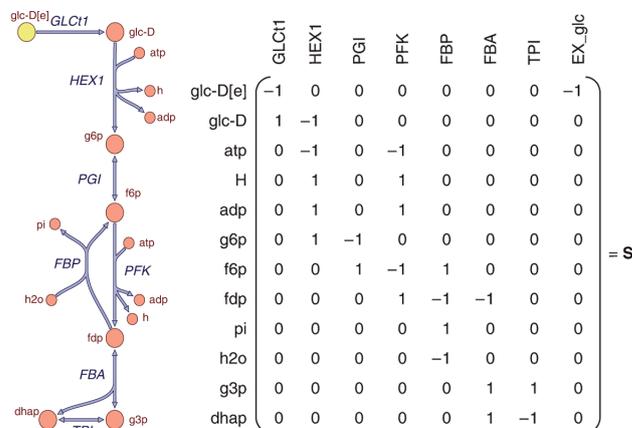


Figura 1.6: A sinistra una catena di reazioni fra metaboliti procede catalizzata da enzimi, a destra la rispettiva matrice stechiometrica che descrive il processo. Da [3].

e cioè:

$$\frac{dx}{dt} = Sv \quad (1.1)$$

Il vettore dei flussi individua nei suoi elementi il tasso di velocità delle singole reazioni, mentre la derivata temporale viene applicata al vettore delle concentrazioni. La matrice stechiometrica trasforma linearmente il primo nel secondo. Alla matrice stechiometrica sono associati quattro sottospazi vettoriali fondamentali e vengono mostrati in figura 1.7. La derivata temporale è nello spazio colonna di S e può essere scritta come espansione di Sv

$$\frac{dx}{dt} = s_1v_1 + s_2v_2 + \dots + s_nv_n$$

dove  $s_i$  sono i vettori reazioni che formano le colonne di S, che sono caratteristiche strutturali del network e come tali sono costanti. Al contrario, i flussi  $v_i$  sono quantità scalari e sono le variabili. Il vettori dello spazio nullo sinistro rappresentano le relazioni di conservazione della massa. Il vettore flusso può essere decomposto in una parte dinamica e una parte stazionaria:

$$v = v_{dyn} + v_{ss}$$

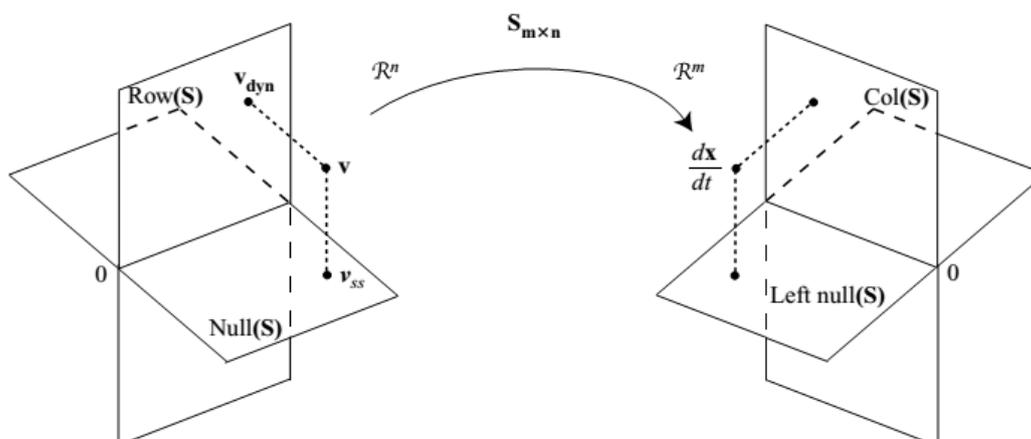


Figura 1.7: La trasformazione lineare  $S$  e i quattro sottospazi fondamentali. Da [4].

La componente statica soddisfa la relazione:

$$Sv_{ss} = 0 \quad (1.2)$$

e quindi  $v_{ss}$  è lo spazio nullo di  $S$ , mentre la componente dinamica del vettore dei flussi è ortogonale allo spazio nullo e come tale è nello spazio righe di  $S$ . I tipi di reazioni biochimiche presenti all'interno della cellula sono in numero limitato e possono essere ridotti a tre categorie rappresentabili da matrici stechiometriche adeguate. Utilizzando una notazione in cui  $C$  indica un metabolita,  $P$  un gruppo fosfato ed  $A$  un cofattore abbiamo:

- conversioni reversibili, del tipo  $CP \rightleftharpoons PC$

$$S = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

- associazioni bi-molecolari, del tipo  $C + P \rightleftharpoons CP$

$$S = \begin{pmatrix} -1 & 1 \\ -1 & 1 \\ 1 & -1 \end{pmatrix}$$

- reazione accoppiata da un cofattore, del tipo  $C + AP \rightleftharpoons CP + A$

$$S = \begin{pmatrix} -1 & 1 \\ -1 & 1 \\ 1 & -1 \\ 1 & -1 \end{pmatrix}$$

Possiamo a questo punto considerare la matrice stechiometrica topologica  $\hat{S}$  così costituita

$$s_{ij} = 0 \rightarrow \hat{s}_{ij} = 0$$

$$s_{ij} \neq 0 \rightarrow \hat{s}_{ij} = 1$$

che rappresenta la forma binaria di  $S$ : se  $\hat{s}_{ij}$  è l'unità, allora il componente  $i$  partecipa alla reazione  $j$ . Sulla base di questa matrice possiamo subito introdurre proprietà topologiche fondamentali. La somma degli elementi non nulli di una colonna

$$\pi_j = \sum_{i=1}^m \hat{s}_{ij}$$

che fornisce il numero di componenti che partecipano alla reazione  $j$ , chiamata numero di partecipazione per la reazione. La somma dei valori unitari in una riga

$$\rho_i = \sum_{j=1}^n \hat{s}_{ij} \quad (1.3)$$

che fornisce il numero di reazioni in cui il composto  $i$  prende parte. Questo numero è una misura di quanto connesso è un composto all'interno del network di reazioni.

## Capitolo 2

# Flux Balance Analysis e Flux Variability

Abbiamo introdotto la matrice stechiometrica che descrive i modelli metabolici e costituisce l'elemento fondamentale di ogni ricostruzione: la forma matriciale consente l'applicazione di tecniche di analisi e perturbazione del modello di tipo matematico, in maniera tale che  $S$  non rappresenti solo un database statico di informazioni metaboliche ma costituisca anche la base per simulazioni ed indagini attive. Fra le tecniche che sono emerse negli ultimi anni con lo scopo di approfondire l'applicazione della matrice stechiometrica  $S$  alla simulazione del metabolismo si possono citare:

- Extreme Pathways analysis [5], in cui lo stato metabolico del sistema è descritto da una base convessa di vettori che rappresentano stati stazionari del network. Per ogni ricostruzione, esiste un particolare set unico di questi vettori. Con l'aggiunta di alcuni vincoli (come conservazione della massa e velocità delle reazioni) è possibile identificare uno spazio di soluzioni, ulteriormente riducibile ad una sola soluzione nel caso sia disponibile un modello cinetico (variazione delle concentrazioni) del sistema.
- Elementary mode analysis [6]: tecnica molto simile a quella degli extreme pathways, anche in questo caso viene selezionato un gruppo ristret-

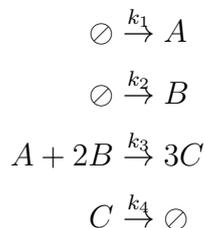
to di vettori per la descrizione dello stato stazionario di un particolare network. Inoltre vengono aggiunti vincoli di tipo termodinamico per valutare se la soluzione sia accettabile o meno.

- Minimal metabolic behaviors [7]: anche in questo caso si tratta di identificare un cono di soluzioni nello spazio vettoriale individuato dal network, utilizzando vincoli non negativi.

In questo lavoro di tesi viene esposta per esteso ed utilizzata una quarta tecnica, conosciuta come Flux balance analysis (da ora in poi FBA).

## 2.1 Ipotesi sul sistema

La FBA è una tecnica basata su vincoli applicabile a modelli stechiometrici del metabolismo: il suo obiettivo è quello di calcolare il flusso di metaboliti attraverso un network di reazioni, al fine di rendere calcolabili parametri del fenotipo cellulare di cui ci si sta occupando. I vincoli sono rappresentati dai coefficienti di  $S$ , i quali impongono flussi sulle reazioni in cui la massa totale dei partecipanti viene conservata. Inoltre ad ogni reazione può essere assegnato un flusso massimo e minimo, riducendo lo spazio delle configurazioni disponibili al sistema. La FBA è applicabile sotto alcune ipotesi: la prima riguarda i parametri cinetici delle reazioni. Consideriamo il seguente sistema di reazioni



dove  $\circlearrowleft$  rappresenta una reazione in cui il metabolita viene prelevato o ceduto dall'esterno della cellula o da un altro compartimento (non rappresentato

dal modello) e i parametri  $k_i$  rappresentano la velocità di reazione. Le corrispondenti equazioni differenziali basate sulla legge della massa descrivono come varia la concentrazione dei metaboliti A-B-C nel tempo sono

$$\begin{aligned}d[A] / d[t] &= k_1 - k_2 [A] [B]^2 \\d[B] / d[t] &= k_2 - 2k_3 [A] [B]^2 \\d[C] / d[t] &= 3k_3 [A] [B]^2 - K_4 [C]\end{aligned}$$

Studiare la dinamica vuol dire conoscere quindi i tassi di velocità di reazione per ognuna di queste equazioni, dati che non sono accessibili per le migliaia di reazioni che compongono una ricostruzione metabolica (soprattutto perché tali misurazioni devono avvenire in vivo). La semplificazione principale della FBA prevede che il lato destro delle equazioni differenziali non preveda dipendenza dalla concentrazione dei metaboliti, mantenendo solo la descrizione stechiometrica dei metaboliti. In tal senso, i flussi passanti per ogni reazione rappresentano l'unica informazione necessaria per studiare il tasso della reazione. Abbiamo

$$\begin{aligned}d[A] / d[t] &= v_1 - v_3 \\d[B] / d[t] &= v_2 - 2v_3 \\d[C] / d[t] &= 3v_3 - v_4\end{aligned}$$

dove  $v = [v_1, v_2, v_3, v_4^T]$  rappresenta il vettore dei flussi per ogni reazione. Come già menzionato, una trasformazione metabolica avviene se è espresso ed attivo l'enzima che la catalizza: ogni reazione di questo tipo è quindi subordinata all'espressione del gene che codifica l'enzima, espressione che avviene su una scala di tempi molto più lunga rispetto al processo metabolico stesso che ne consegue. La trascrizione di un singolo gene e lo spostamento in sito della proteina richiedono 160s nel batterio di E.Coli, mentre il processo metabolico che ne consegue richiede un tempo dell'ordine dei 60ms [8].

Questo vuol dire che nei tempi di interesse, quelli metabolici, la cellula ha già ampiamente raggiunto lo stato stazionario in cui la derivata delle concentrazioni dei metaboliti è nulla. Possiamo riscrivere l'equazione di prima con l'aggiunta di questa ipotesi in forma matriciale.

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 1 & 0 & -2 & 0 \\ 0 & 0 & 3 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

In forma compatta possiamo scrivere per ogni matrice  $S$  nello stato stazionario appena illustrato:

$$Sv = 0 \quad (2.1)$$

In aggiunta a queste ipotesi possiamo introdurre dei limiti superiori ed inferiori ai flussi di metaboliti per ogni reazione. Questo tipo di vincoli possono avere una natura sperimentale se sono stati effettivamente misurati o possono essere decisi ad hoc per la simulazione che vogliamo operare. I flussi avranno valori positivi se vanno nel verso dei prodotti, negativi se la reazione viene percorsa al contrario nel senso dei reagenti.

$$L_i \leq v_i \leq U_i, \quad \forall i = 1, \dots, r. \quad (2.2)$$

Dove  $L_i$  rappresenta il flusso minimo e  $U_i$  quello massimo per ogni reazione. Nel caso in cui  $L_i = 0$  o  $U_i = 0$  la reazione  $i$  è irreversibile e può essere percorsa in un solo senso. I vincoli introdotti generano un insieme convesso  $W \in \mathfrak{R}$  a cui i flussi  $v$  devono appartenere: per ottenere il punto in  $W$  che descrive l'effettiva configurazione dei flussi nel network è necessario un processo di ottimizzazione da basarsi sull'introduzione di una funzione obiettivo da massimizzare (o minimizzare). Il metabolismo è un meccanismo con il quale nutrienti sono convertiti in energia e nei blocchi per costituire macromolecole (amminoacidi, proteine..): questi ultimi sono i costituenti della massa cellulare e sono necessari affinché la cellula si duplichi e prolifichi. Nel loro insieme questi blocchi formano la biomassa della cellula: le quantità e proporzioni di questi costituenti possono essere misurati sperimentalmente.

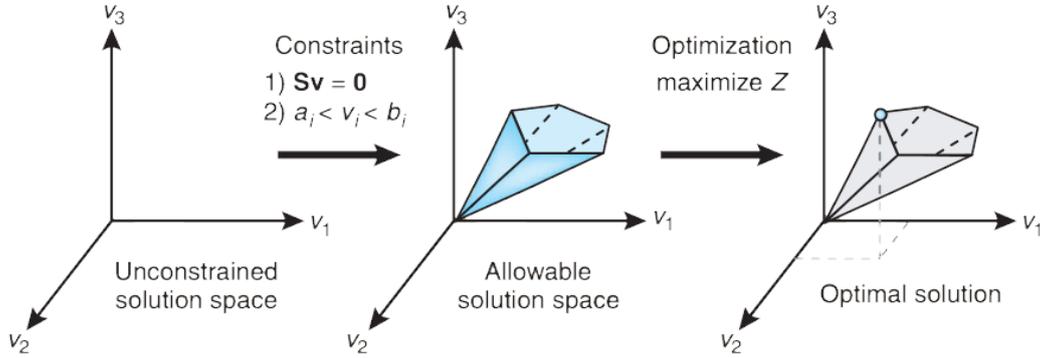


Figura 2.1: a) Lo spazio senza vincoli dei flussi. b) L'introduzione dei vincoli crea un cono convesso di soluzioni possibili. c) La massimizzazione della funzione obiettivo ( $Z$  nell'immagine) converge ad un solo risultato ottimale. Da [10].

Si decide di assumere che l'obiettivo del metabolismo sia quello di massimizzare la biomassa e che quindi l'evoluzione abbia selezionato quelle reazioni e quei percorsi cellulari atti alla produzione di quanta più materia cellulare possibile. In questo senso la produzione di biomassa diventa la funzione di obiettivo  $\phi(v)$  definita come combinazione lineare di flussi che descrivono i precursori della biosintesi che ci interessa:

$$\phi(v) = b^T v \quad (2.3)$$

dove  $b$  contiene i coefficienti ottenuti dalle misurazioni sperimentali della biomassa, cioè la quantità di precursori necessari ad ottenere un'unità di biomassa. Nel caso di cellula non perturbata, i flussi metabolici possono essere ricavati secondo questa ottimizzazione:

$$v^* = \arg \max_{w \in W} b^T w \quad (2.4)$$

dove  $\arg \max_{c \in Q} f(x)$  indica quel set di  $x$  per il quale la funzione  $f$  ottiene il massimo valore in  $Q$ . Quando una cellula è perturbata (ad esempio in seguito ad una modifica alla matrice  $S$  che la rappresenta) essa si allontana dallo stato stazionario 2.1 ed è necessario introdurre un nuovo criterio da applicare al

processo di ottimizzazione. Il più utilizzato in questo senso è il MOMA: secondo [9] a seguito di una perturbazione la cellula reagisce minimizzando gli aggiustamenti rispetto alla situazione non perturbata di 2.4. Il problema di ottimizzazione che si andrà a risolvere è il seguente:

$$v^{MOMA} = \arg \min_{w \in W} \|w - v^*\|^2 \quad (2.5)$$

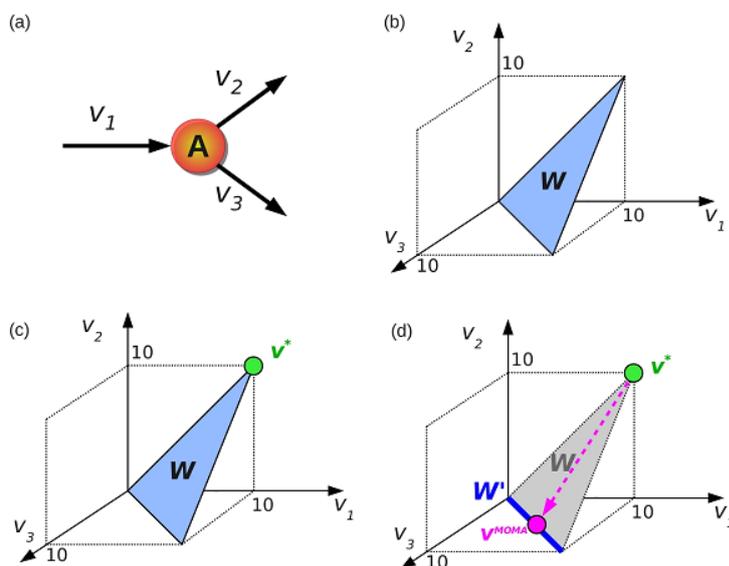


Figura 2.2: a) Semplice network in cui il metabolita A è prodotto da  $v_i$  e consumato da  $v_2$  e  $v_3$ . b) Set convesso di soluzioni possibili nello spazio dei flussi, definito dal vincolo di conservazione  $v_1 - v_2 - v_3 = 0$  e  $\max(v_i) = 0$ . c)  $v^*$  rappresenta il punto individuato dall'ottimizzazione secondo la funzione obiettivo  $b = [0, 2, 1]^T$ . d) Il network viene perturbato dalla rimozione di  $v_2$ , il set di soluzioni si riduce da  $W$  a  $W'$  e per la scelta del criterio MOMA il nuovo risultato  $v^{MOMA}$  è la proiezione sul nuovo spazio del risultato non perturbato  $v^*$ . Da [2].

## 2.2 Programmazione Lineare

L'ultimo step FBA è un problema di programmazione lineare. Con questo termine si indica quella branca della ricerca operativa indirizzata alla risoluzione di problemi di ottimizzazione lineare: un problema si definisce lineare quando sia la funzione obiettivo che i vincoli imposti al sistema sono di tipo lineare ed in questo caso si parla di ottimizzazione convessa. Il modo standard di presentare un problema lineare è scomponendolo in tre parti ( $v \in \mathbb{R}^r$ )

- vincoli: in quanto lineari possono essere scritti in genere come un insieme di uguaglianze

$$A_{eq}v = c_{eq}$$

e disuguaglianze

$$A_{ineq}v \leq c_{ineq}$$

quindi senza perdere di generalità possiamo scrivere per entrambi

$$Av \leq c$$

- variabili non negative: nella forma standard vengono considerati solo valori non negativi per le variabili

$$v_i \geq 0 \quad \forall_i = 1, \dots, r$$

- funzione obiettivo  $\phi$ : è definita come il prodotto fra il vettore dei flussi e un vettore  $b$

$$\phi(v) = b^T v$$

Di estrema importanza è il risultato per cui un problema di ottimizzazione così costituito o non ha nessun risultato o ne ha uno ed uno soltanto o ne ha infiniti. La forma appena esposta prende il nome di problema primario: per ogni problema di ottimizzazione primario esiste il suo duale che fornisce un

limite superiore alla soluzione del primo. In forma matriciale, se il problema primario è esposto come segue

$$\text{Massimizzare } b^T v \text{ soggetto a } Av \leq c, v \geq 0$$

il problema duale è esprimibile come

$$\text{Minimizzare } c^T w \text{ soggetto a } A^T w \geq b, w \geq 0$$

dove  $w$  è il vettore duale delle variabili. L'importanza del problema duale risiede nel teorema duale forte della programmazione lineare:

**Theorem 1.** *Sia  $A$  una matrice e siano  $b$  e  $c$  vettori. Allora*  
 $\max\{b^T v \text{ tale che } Av \leq c, v \geq 0\} = \min\{c^T w \text{ tale che } A^T w \geq b, w \geq 0\}$   
*posto che  $w$  e  $v$  non siano vuoti.*

Questo significa che all'ottimo, la funzione obiettivo primaria e duale assumono lo stesso valore, cosicché è possibile imporre l'uguaglianza fra i due:

$$b^T v = c^T w$$

Questa relazione è alla base dei sistemi risolutivi implementati dai pacchetti software preposti all'ottimizzazione di problemi lineari al computer.

## 2.3 Applicazioni e limiti del FBA: la Flux Variability Analysis

La FBA ha trovato fin'ora applicazioni nell'ingegneria dei bioprocessi, dove è stato utilizzata per identificare modifiche al network metabolico di microbi capaci di incrementare alcuni prodotti derivanti dal processo di fermentazione [11]. E' stata utilizzata per individuare combinazioni multidrugs per la cura del cancro ad impatto minimo [2], per lo studio razionalizzato di culture di crescita [12] e nello studio di interazione fra ospiti ed organismi patogeni [13]. Il risultato di un analisi FBA può essere visualizzato usando

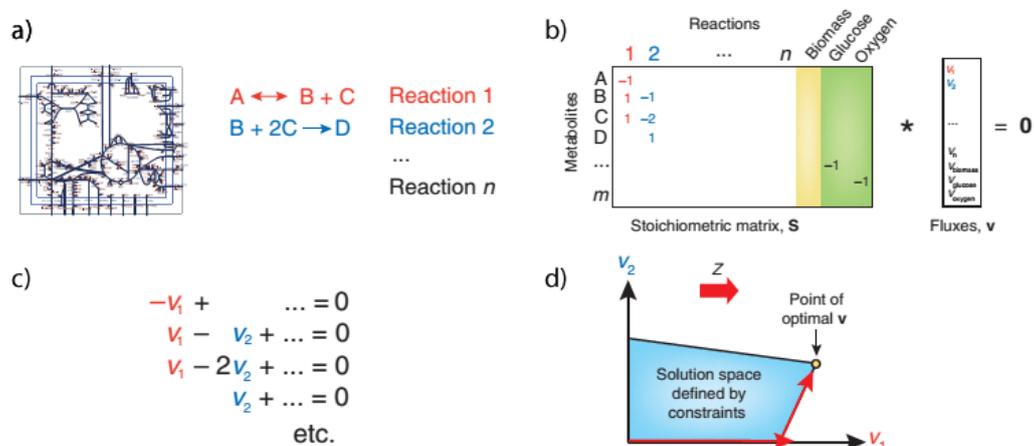


Figura 2.3: a) Ricostruzione del network metabolico b) Matrice  $S$  e condizione di stato stazionario c) La conservazione della massa definisce un sistema di equazioni d) Dopo aver definito  $Z$  la funzione obiettivo da massimare, la FBA individua la soluzione. Da [10].

mappe di flusso come nell'immagine 2.4. Un inconveniente che si può incontrare in ogni ottimizzazione basata sui vincoli è la possibile esistenza di soluzioni ottimali alternative: questi risultati alternativi sono costituiti da set di reazioni con diverse distribuzioni di flusso che possono essere usate dal sistema per raggiungere lo stesso valore della funzione obiettivo che viene massimizzata. L'impatto di questa situazione sulle conclusioni ottenute a partire da simulazioni di flusso possono essere trascurabili o di elevata importanza a seconda dei casi ed è bene essere consapevoli nel momento in cui si decide di utilizzare la FBA. Come verrà esposto, questa proprietà emerge in maniera evidente se si prova a risolvere lo stesso problema di ottimizzazione con più di un risolutore numerico al calcolatore: a seconda del metodo implementato da ciascuno dei risolutori la distribuzione dei flussi potrebbe essere anche molto diversa, pur conservando il medesimo risultato finale. Tanto più la ricostruzione è poco vincolata da estremi di flusso e relazioni stechiometriche, tanto più lo spettro di soluzioni sarà ampio. Questa situazione, che trova anche senso biologico nella ridondanza dei pathways, può essere studia-

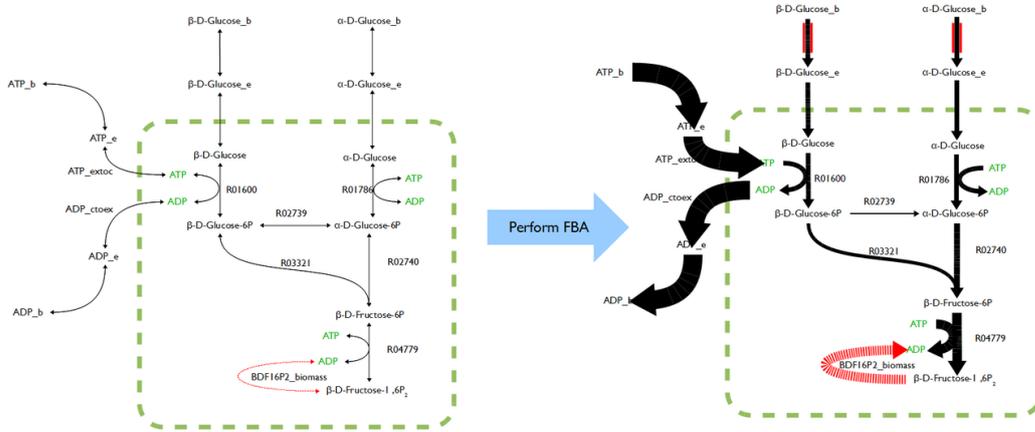


Figura 2.4: Visualizzazione di un risultato di FBA, flussi stazionari delle reazioni della glicolisi. Lo spessore delle frecce è proporzionale al flusso lungo la reazione. Da wikipedia.org.

ta se si ricorre ad una versione modificata della FBA, ossia la flux variability analysis [14] (d'ora in poi FVA). La FVA è usata per trovare il minimo e il massimo di ogni flusso per quelle reazioni che appartengono al set delle soluzioni alternative trovate dalle FBA: mantenendo quindi inalterata la funzione obiettivo e il valore di ottimo, la FVA trova gli estremi di flusso che portano a quel risultato, eliminando l'ambiguità della FBA. Inoltre è generalmente possibile intervenire sui parametri dell'ottimizzazione in maniera da trovare quei flussi che portano ad un risultato perfettamente equivalente all'ottimo o ad una percentuale di esso. Dopo aver risolto il problema FBA, l'FVA risolve due problemi di ottimizzazione per ogni flusso  $v_i$

$$\begin{aligned}
 & \max_v / \min_v \quad v_i \\
 & \text{subject to} \quad Sv = 0 \\
 & \quad \quad \quad b^T v \geq \gamma Z_0 \\
 & \quad \quad \quad v_l \leq v \leq v_u
 \end{aligned}$$

dove  $Z_0 = b^T v_0$  è l' ottimo del problema FBA e  $\gamma$  è un parametro con il quale controllare la percentuale che si vuole ottenere del risultato massimo della

funzione obiettivo ( $\gamma < 1$  stato sub ottimale). Assumendo tutte le  $n$  reazioni di un network metabolico interessanti dal punto di vista della FVA, sarà necessario risolvere  $2n$  problemi lineari. Applicazioni della FVA riguardano lo studio di ottimi alternativi per la funzione obiettivo [14], studiare la crescita cellulare (biomassa) in condizioni sub-ottimali [15], investigare la flessibilità del network metabolico e la sua ridondanza [16] etc ...



# Capitolo 3

## Recon 2

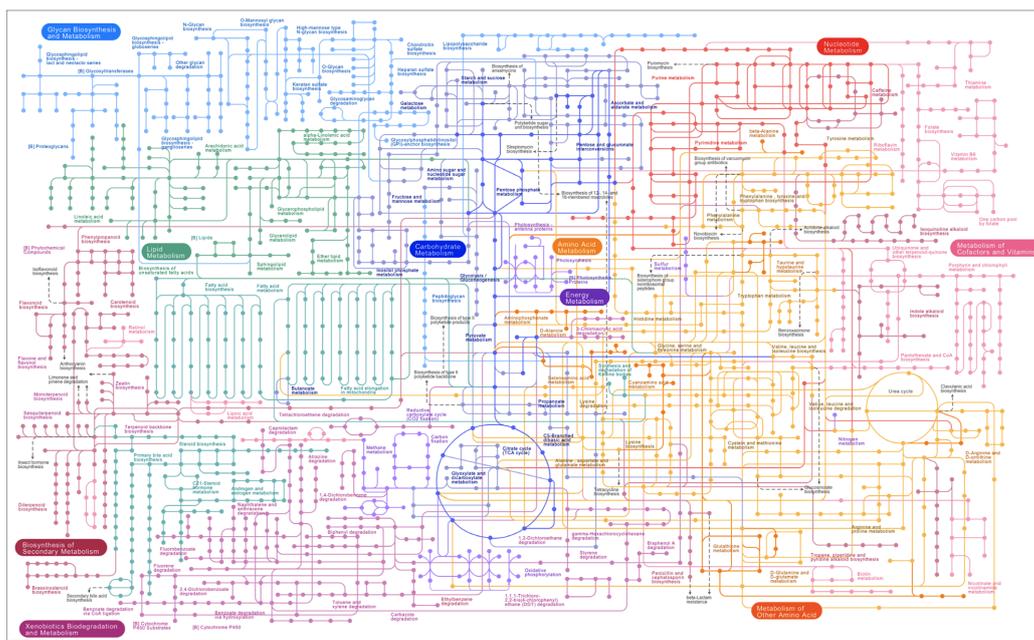


Figura 3.1: Esempio di ricostruzione del metabolismo umano: mappa dei pathways catalogati dal database genomico KEGG [17].

Sono state fino ad ora esposte le basi necessarie ad introdurre i metodi ed i materiali utilizzati in questo lavoro di tesi: tutta l'analisi che verrà presentata nei successivi capitoli si è svolta sulla ricostruzione del metabolismo umano

più completa attualmente disponibile, Recon2 [18]. Questa ricostruzione è nata come espansione del precedente Recon1 [19] a seguito di incontri avvenuti fra le comunità di ricerca interessate alle ricostruzioni metaboliche e vanta circa il doppio di reazioni e metaboliti della prima versione. In effetti Recon2

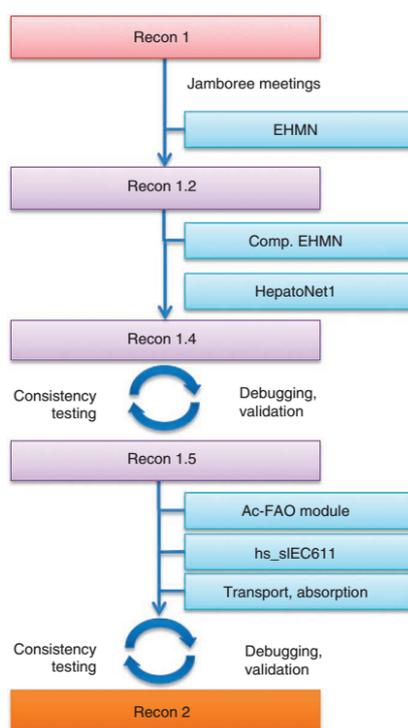


Figura 3.2: Le varie fasi dello sviluppo di Recon2: a partire da Recon1 ed in seguito agli incontri fra le comunità sono stati integrati nuovi database metabolici (in azzurro). Dopo un certo numero di modifiche il modello è stato testato prima di procedere alle successive e definitive inclusioni. Da [18]

conta in aggiunta alla prima versione l'integrazione con altre 4 ricostruzioni (EHMN, HepatoNet1, Ac/FAO module e la ricostruzione metabolica di una cellula epiteliale dell'intestino) e le informazioni in esso contenute sono annotate secondo standard semantici. E' possibile scaricare liberamente Recon2 dal sito <http://humanmetabolism.org/>.

### 3.1 Le caratteristiche della ricostruzione

Una volta scaricata l'ultima versione di Recon2 dal sito web del gruppo di ricerca, è possibile importare il modello direttamente in Matlab [20] in quanto file di estensione .mat: la versione fornita si presenta come variabile di tipo struct, a cui è possibile accedere ai vari sottocampi. Nella tabella

Campo	Dimensione	Descrizione
S	5063x7440	matrice stechiometrica
rxns	7440	nome abbreviato reazioni
lb	7440	vincolo inferiore di flusso per la reazione
ub	7400	vincolo superiore di flusso per la reazione
rev	7440	reversibilità della reazione
c	7440	funzione obiettivo
GenesMat	7440x2194	associazione reazione - geni
genes	2194	geni
subSystems	7440	pathway d'appartenenza della reazione
rxnECNumbers	7440	identificativo EC della reazione
mets	5063	nome abbreviato metabolita
metNames	5063	nome esteso metabolita
metKeggID	5063	identificativo Kegg del metabolita

Tabella 3.1: Alcuni dei campi presenti nella struttura matlab di Recon2

3.1 vengono mostrati i campi di Recon2 che verranno presi in considerazione in questo lavoro, con le loro rispettive dimensioni e breve descrizione. In totale, i campi di Recon2 sono 36. La matrice stechiometrica S è composta da 31.540 valori non nulli, con una sparsità dello 0.0837%. Circa 28.700 dei valori non nulli sono rappresentati da 1 o -1. Il valore massimo di S risulta essere -18244, coefficiente stechiometrico attribuito ai metaboliti H<sub>2</sub>O e ATP nella reazione 'apolipoprotein B Protein assembly'. Infine, 261 valori non sono numeri interi. Ad ogni reazione è associato un pathway di riferimento,

segnato nel campo `subSystems`, in cui questa si svolge: il numero di pathway unico è 100, fra cui un generico ‘non-assigned’ per quelle reazioni di cui non si conosce il pathway di appartenenza. Quando possibile, ad ogni reazione nota è associato un identificato EC, ossia l’enzyme Commission number che identifica univocamente il processo catalizzato dall’enzima [22]. Il numero di metaboliti unici è 2957, ma questi vengono suddivisi e differenziati nel nome a seconda del compartimento cellulare in cui si trovano: esistono 8 compartimenti diversi con relativi suffissi, portando il numero totale dei metaboliti a 5063. I compartimenti sono: spazio extra cellulare (suffisso [e]), citoplasma ([c]), mitocrondri ([m]), nucleo ([n]), reticolo endoplasmatico ([r]), perossisoma ([x]), lisosoma ([l]) e golgi ([g]). Ad ogni metabolita è attribuito anche un codice identificativo univoco Kegg. I vincoli `lb` e `ub` rappresentano i flussi massimi per ogni reazione di recon, nelle due direzioni di percorrenza possibili per la reazione: `ub` sarà il flusso massimo nella direzione dei prodotti mentre `lb`, di segno opposto, rappresenta il flusso massimo in direzione dei reagenti. Così come viene fornita la ricostruzione, i flussi in valore assoluto si distribuiscono fra 0, 1, 100 e 1000. La colonna `rev` è di tipo binario e riporta valore unitario per quelle reazioni che biologicamente possono essere percorse in entrambi i sensi: il 50.2% delle reazioni di recon sono segnalate come reversibili. Il campo `c` è una colonna anch’essa binaria che identifica con il valore 1 la reazione (o più reazioni) selezionata come funzione obiettivo di un’eventuale simulazione di FBA. Come verrà illustrato, nel caso di Recon2 la colonna `C` assume valore 1 in corrispondenza della reazione biomassa umana. Nella ricostruzione di recon sono stati ovviamente inseriti anche i geni metabolici che si suppone codifichino enzimi per una o più delle 7440 reazioni tabulate: la matrice `GenesMat` associa per ogni reazione i geni associati, mentre in `genes` troviamo i soli codici dei geni in nomenclatura perlopiù `entrez` [21]. I campi di recon che non sono stati riportati in questa sede riguardano principalmente nomenclature alternative (`CHEBIID`, `PubChemID`, `HepatoNetID`...) o dati non strettamente necessari per la nostra analisi come carica chimica dei metaboliti e la loro formula chimica completa.

## 3.2 Reazione di biomassa umana

Merita un approfondimento a parte la reazione di biomassa che, come esposto nella sezione 3.2, costituisce la funzione obiettivo delle simulazione di FBA e FVA. Il vettore binario  $c$  di Recon2 indica con l'unico valore unitario nella sua colonna la posizione della funzione obiettivo, che è la numero 6799 e viene riportata come 'Generic Human Biomass Reaction'. Tale reazione in realtà non esiste come unico processo all'interno della cellula, ma racchiude tutti quei mattoni essenziali affinché la cellula prolifichi e si replichi: utilizzando però questa reazione fittizia come funzione obiettivo, la FBA massimizza tutti quei processi che hanno come esito finale la produzione di uno o più dei metaboliti che compongono la biomassa. La tabella 3.2 riporta la composizione di tale reazione: il numero di metaboliti reagenti è 38 ed i prodotti 3, ed il pathways di riferimento per questo processo come riportato dal campo subSystems di recon è 'exchange/demand reaction'. Tutti i metaboliti partecipanti a questa reazione si trovano nel citoplasma, a parte i precursori dei nucleotidi (dGTP, dCTP, dATP, dTTP) che si trovano ovviamente all'interno del nucleo. Il processo di formulazione della reazione di biomassa comincia con la definizione del contenuto macromolecolare della cellula (frazione di peso di proteine, RNA, lipidi, etc..) e quindi dei metaboliti che competono a ciascun gruppo (aminoacidi, nucleotidi..). Con queste informazione è quindi possibile dettagliare le quantità richieste di metaboliti che sono necessari, con i relativi pathways. E' possibile incrementare il livello di risoluzione e calcolare anche la quantità di energia necessaria alla sintesi di macromolecole appena evidenziate. Per esempio, è noto che sono necessarie approssimativamente 2 molecole di ATP e 2 di GTP per completare la polimerizzazione di ogni aminoacido [23] in una proteina. Questa energia di conversione è inclusa nella reazione di biomassa, che include tutta l'energia necessaria affinché la cellula concluda i processi biosintetici. Un ulteriore passo consiste nell'aggiunta dettagliata di vitamine, elementi e co-fattori necessarie alla crescita. Questo tipo di considerazioni sono alla base della reazione di biomassa umana: i coefficienti stechiometrici per ogni me-

Metabolita	Coeff.	Metabolita	Coeff.
Acqua	20.70	Cardiolipin	0.012
ATP	20.70	UTP	0.053
L-Glutamate	0.38	dGTP	0.01
L-Aspartate	0.35	dCTP	0.009
L-Alanine	0.28	D-Glucose-6-Phosphate	0.27
L-cysteine	0.046	L-histidine	0.13
L-glutamine	0.32	L-tyrosine	0.16
Glycine	0.54	L-isoleucine	0.29
L-serine	0.39	L-leucine	0.54
L-threonine	0.31	L-tryptophan	0.13
L-argininium	0.36	L-proline	0.41
L-methionine	0.15	Phosphatidylserine	0.006
1-phosphatidyl-1D-myo-inositol	0.02	Sphingomyelin betaine	0.017
CTP	0.04	L-valine	0.35
Phosphatidylcholine	0.15	<b>APD</b>	20.65
Phosphatidylethanolamine	0.055	<b>Proton</b>	20.65
Cholesterol	0.02	<b>Idrogenofosfato</b>	20.65
Phosphatidylglycerol	0.003		

Tabella 3.2: Composizione della reazione di biomassa umana: metaboliti con i loro coefficienti stechiometrici. In grassetto i 3 metaboliti prodotti dalla reazione.

tabolita sono effettivamente il frutto di considerazioni di tipo energetico e sperimentale, in cui le frazioni di macromolecole sono state misurate e riportate in maniera stechiometricamente proporzionata nella reazione [24]. La composizione macromolecolare è così costituita: proteine 70.6%, DNA 1.4%, RNA 5.8%, carboidrati 7.1%, lipidi 9.7%, altri 5.4%: il carboidrato principale è il glucosio-6-fosfato. In ‘altri’ sono contenute le vitamine, piccole molecole, ioni, cofattori etc..In totale, il peso cellulare assunto per questa distribuzione è di  $500 \times 10^{-12}$  grammi.

### 3.3 La rete dei metaboliti e delle reazioni

Come è stato illustrato, la matrice  $S$  di una ricostruzione metabolica mette in relazione i metaboliti con le reazioni in cui questi partecipano: in questo senso la matrice  $S$  è una matrice d’incidenza, dove ad ogni reazione corrisponde una colonna e ad ogni riga un metabolita. La struttura di  $S$  inoltre è matematicamente pronta per essere utilizzata in simulazione di FBA, FVA e così via. Potrebbe essere interessante però sviluppare le relazioni esistenti fra i rispettivi elementi della matrice stechiometrica separatamente: a tal fine è necessario estrarre da  $S$  due network distinti, i cui nodi sono i metaboliti o le reazioni ed i link i legami fra questi elementi. Nel caso della rete metaboliti-metaboliti è possibile estrarre da  $S$  il network che cerchiamo applicando il criterio esposto nell’immagine 3.3: ogni volta che un metabolita è prodotto da una reazione, questo è collegato ai metaboliti reagenti che hanno concorso alla sua formazione. La rete che si viene a creare è diretta (i link vanno dai prodotti ai reagenti) e pesata dal numero di volte in cui uno stesso prodotto è stato formato da reagenti con lo stesso metabolita. Nella costruzione di questo network viene considerata anche la colonna rev della reversibilità delle reazioni: se una reazione può essere percorsa nel senso opposto, il ruolo reagente-prodotto si inverte, rendendo possibili nuovi legami nella logica di collegamento che è appena stata descritta. Un discorso analogo si può applicare al caso della rete delle reazioni, come mostrato nell’immagine

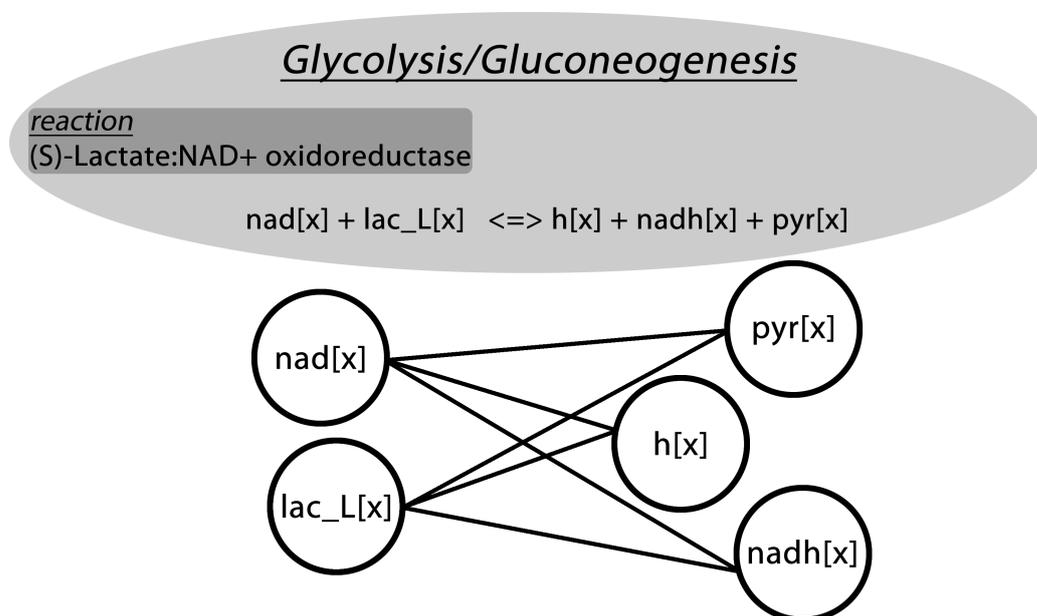


Figura 3.3: Criterio per la creazione dei link nella rete dei metaboliti

3.4. In questo caso il criterio adottato è il seguente: nel network reazione-reazione, un link fra due reazioni si viene a creare ogni volta il prodotto di una reazione è reagente in un altro. Anche in questo caso la rete è diretta ed è pesata dal numero di metaboliti che una reazione fornisce come reagenti ad altra. La reversibilità viene di nuovo presa in considerazione per includere tutte quei processi che possono essere percorsi in senso opposto. Abbiamo ottenuto in questa maniera due network distinti, che catturano le relazioni esistenti fra i metaboliti e le reazioni fra loro, separatamente.

### 3.4 Misure di centralità

Abbiamo ottenuto due network: il primo è quello metaboliti-metaboliti, di dimensione 5063x5063 e sparsità 0.066%, che cattura i legami esistenti fra gli elementi che compongono le reazioni. Il secondo network è quello reazioni-reazioni, di dimensione 7440x7440 e sparsità 3.06%, che racchiude le relazioni fra le reazioni di S. Possiamo indagare questi network applicando semplici

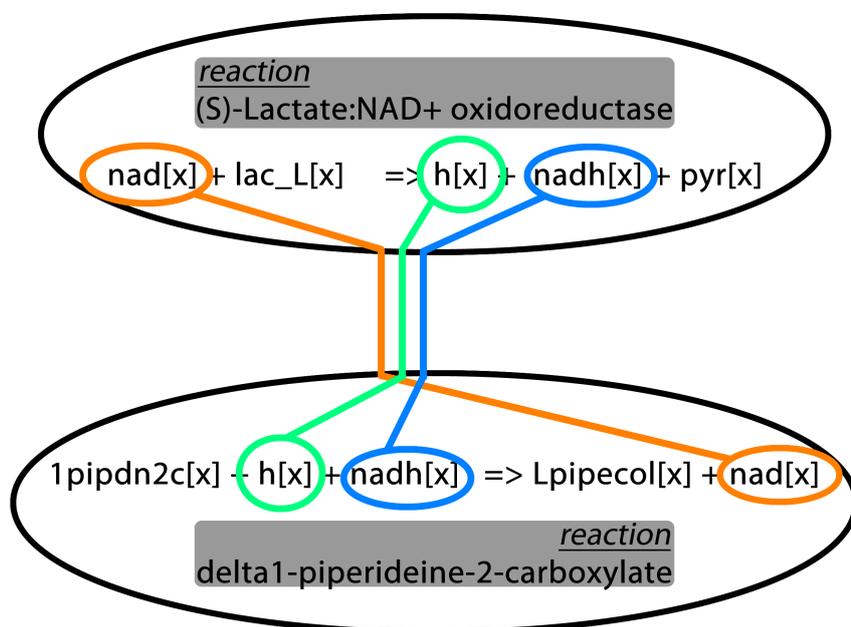


Figura 3.4: Criterio per la creazione dei link nella rete delle reazioni: in base ai metaboliti prodotti e reagenti la seconda reazione è legata alla prima con un link di valore 2, la prima è legata alla seconda con un link di valore 1.

misure di centralità. Cominciamo con la connettività dei nodi definita come in 1.3: dopo aver simmetrizzato le matrici (perdendo così informazioni sulla direzionalità dei legami ma guadagnando in generalità), possiamo binarizzare i valori per procedere al conteggio dei link di ciascun nodo dei networks. Guardando il risultato in immagine 3.5, 3.6 si nota come la maggior parte dei nodi (siano metaboliti, che reazioni) hanno un numero basso di connessioni. Come però è possibile valutare dalle tabelle 3.4 e 3.4, esistono alcuni nodi outsider la cui connettività è molto più alta. Nel caso dei metaboliti, questi nodi sono elementi comuni come idrogeno, acqua, ATP e nadp. Anche nel caso delle reazioni i nodi con connettività più alta sono quelli che rappresentano processi che si svolgono fra metaboliti ubiquitari all'interno della cellula, che vengono quindi 'scambiati' da molte reazioni. Più interessante è prendere in considerazione un'altra misura di centralità, ossia la

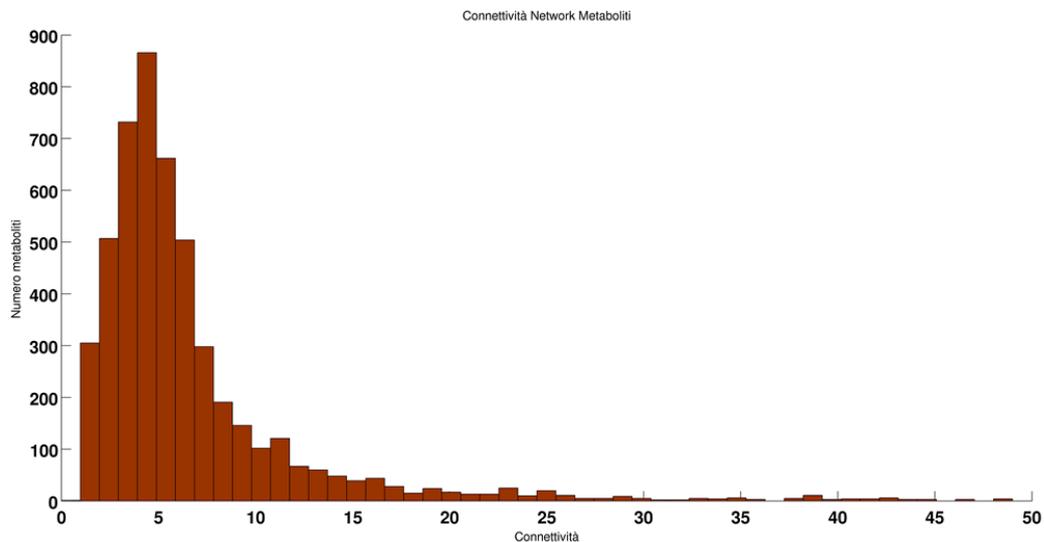


Figura 3.5: Distribuzione della connettività nel network metabolita-metabolita. Viene mostrata la connettività fino a 50 link per nodo.

Metabolita	Compartimento	Conn.
Protone	citoplasma	1185
Acqua	citoplasma	941
Protone	mitocrondrio	450
ATP	citoplasma	415
NADPH	citoplasma	373
NADP	citoplasma	353
Acqua	mitocrondrio	324
Protone	reticolo endoplasmatico	318
Idrogeno fosfato	citoplasma	311
ADP	citoplasma	281

Tabella 3.3: I dieci metaboliti a connettività più alta

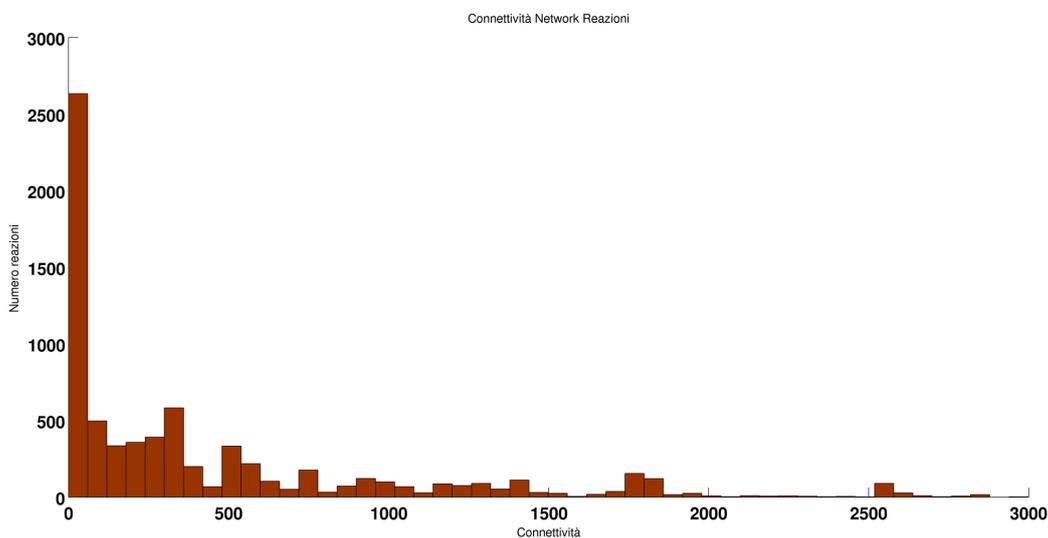


Figura 3.6: Distribuzione della connettività nel network reazione-reazione. Viene mostrata la connettività fino a 3000 link per nodo.

Reazione	Pathway	Conn.
bicarbonate transport	Transport, extracellular	3208
Glutathione transport into mitochondria	Glutathione metabolism	3187
carboxylic acid dissociation	Miscellaneous	3076
transport of palmitoylcoa into peroxisomes	Transport, peroxisomal	3018
bilirubin beta-diglucuronide transport	Transport, extracellular	3011
phosphatidylcholine transporter	Transport, extracellular	3011
bilirubin monoglucuronide transport	Transport, extracellular	3010
transport of Stearoyl-CoA	Transport, peroxisomal	3006
ATP-binding Cassette	Transport, peroxisomal	2998
ATP-binding Cassette (r2518)	Transport, peroxisomal	2998

Tabella 3.4: Le dieci reazioni a connettività più alta

betweenness centrality (d'ora in poi BC) la quale, come suggerisce il nome, è un indicatore di quanto un nodo sia centrale all'interno di un network. La sua definizione per un nodo  $v$  è la seguente:

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

dove  $\sigma_{st}$  rappresenta il numero di shortest paths dal nodo  $s$  al nodo  $t$  e  $\sigma_{st}(v)$  è il numero di quelle passanti per il nodo  $v$ . Vale la pena ricordare che con shortest paths si intende il più breve percorso che collega due nodi di un network, quindi il minor numero di 'passaggi' per passare da un nodo ad un altro. Partendo dall'ipotesi ragionevole per cui anche all'interno di un network di natura biologia il tragitto più breve è quello preferenziale al fine di produrre un metabolita, un nodo con alta BC è di grande importanza perché al centro di molti processi di trasferimento. Per quanto riguarda i metaboliti, il risultato non si discosta molto in termini di composti (protone, acqua, atp) da quanto visto con i risultati di connettività, ma si trova che i compartimenti dove questi rivestono un ruolo più importante vengono suddivisi più equamente fra citoplasma, mitocondri e reticolo endoplasmatico. Più interessante è invece il risultato relativo alle reazioni, dove infatti compare fra le prime dieci la reazione di biomassa della cellula umana, usata da recon come funzione obiettivo per le massimizzazioni FBA. Questo risultato è coerente con l'importanza della funzione e il numero di metaboliti fondamentali che partecipano in essa.

Metabolita	Compartimento	BC
Protone	Citoplasma	1
Acqua	Citoplasma	0.6
Protone	Golgi	0.21
Protone	Reticolo endoplasmatico	0.21
Protone	Mitocondrio	0.16
Acqua	Mitocondrio	0.13
Coenzima A	Citoplasma	0.13
ATP	Citoplasma	0.13
Acqua	Reticolo endoplasmatico	0.13
Protone	Perossisoma	0.12

Tabella 3.5: I dieci metaboliti a betweenness centrality più alta, normalizzata rispetto il valore della prima

Reazione	Pathway	BC
H transporter, endoplasmic reticulum sodium proton antiporter	Transport, endoplasmic reticular	1.00
carbamoyl-phosphate synthase	Transport, golgi apparatus	0.92
H <sub>2</sub> O transport, lysosomal	Pyrimidine synthesis	0.59
potassium transport via proton antiport	Transport, lysosomal	0.55
proton diffusion (Golgi)	Transport, golgi apparatus	0.55
H <sub>2</sub> O endoplasmic reticulum transport	Transport, golgi apparatus	0.55
acetyl-CoA carboxylase	Transport, endoplasmic reticular	0.45
Generic human biomass reaction	Fatty acid oxidation	0.44
H <sub>2</sub> O transport, peroxisomal	Exchange/demand reaction	0.38
	Transport, peroxisomal	0.28

Tabella 3.6: Le dieci reazioni a betweenness centrality più alta, normalizzata rispetto il valore della prima



## Capitolo 4

# Il network dei pathways

Abbiamo visto che la matrice stechiometrica  $S$  racchiude l'intero network metabolico umano, contenendo più di cinquemila metaboliti e più di settemila reazioni: nonostante questi numeri siano già di una certa entità, esiste generale consenso riguardo il fatto possano essere un decimo di quelli effettivi. Per ridurre le dimensioni della ricostruzione senza perdere di vista le caratteristiche di rilievo è necessario spostarsi ad un livello più alto rispetto quello delle reazioni e sfruttare l'organizzazione a pathways del metabolismo, struttura che viene riportata anche nel modello ricostruito. Una versione più compatta del network consente infatti di visualizzare con immagini le strutture metaboliche e individuare immediatamente le connessioni esistenti fra i vari processi che si svolgono nella cellula. A questo scopo viene ora illustrato un possibile modo per sviluppare un network pathway-pathway sulla stregua di quanto già visto per i networks reazione-reazione e metabolita-metabolita 3.3. La rete così costituita ha dimensione  $100 \times 100$ , consentendone per intero la visualizzazione (secondo i criteri che verranno esposti) a differenza dei networks visti fin'ora. Inoltre la natura di questa rete suggerisce anche l'applicazione di modelli markoviani, che posso essere affiancati a quelli già esposti di FBA e FVA per simulazioni e analisi del metabolismo.

## 4.1 Il network pathway-pathway

Come discusso in 3.1 la colonna subSystems della struttura di Recon2 fornisce per ognuna della 7440 reazioni il pathway di appartenenza. In totale esistono 100 pathways fra cui le reazioni si suddividono. In appendice A, la tabella A.1 riporta i 100 pathways e per ciascuno di essi il numero di reazioni e di metaboliti unici che li compongono. Emerge che il gruppo più numeroso di reazioni appartiene alle funzione di trasporto extracellulare, che sono quelle adibite allo scambio di metaboliti con l'esterno della cellula. Come

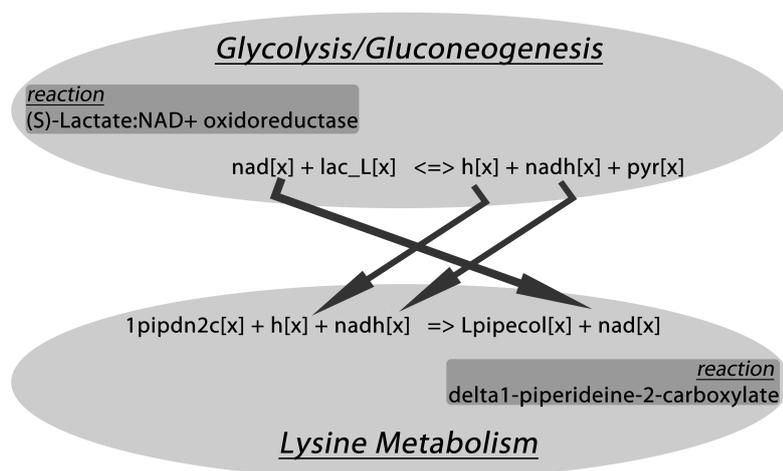


Figura 4.1: Criterio per la creazione dei link nella rete dei pathway: in questo caso il pathways della glicolisi avrà almeno un link di peso due diretto al pathways del metabolismo delle lisine, mentre quest'ultimo avrà almeno un link di peso uno diretto al pathway della glicolisi.

nel caso dei network metabolita-metabolita e reazione-reazione, è necessario introdurre un criterio con il quale estrarre il network pathway-pathway: tutte le volte in cui un metabolita prodotto da una reazione del pathway 'A' è reagente di un pathway 'B', allora esiste un link A-B. Per ogni altro metabolita nella stessa condizione, il link A-B viene incrementato. In questo modo otteniamo il network con matrice di adiacenza 100x100, diretto e pesato con

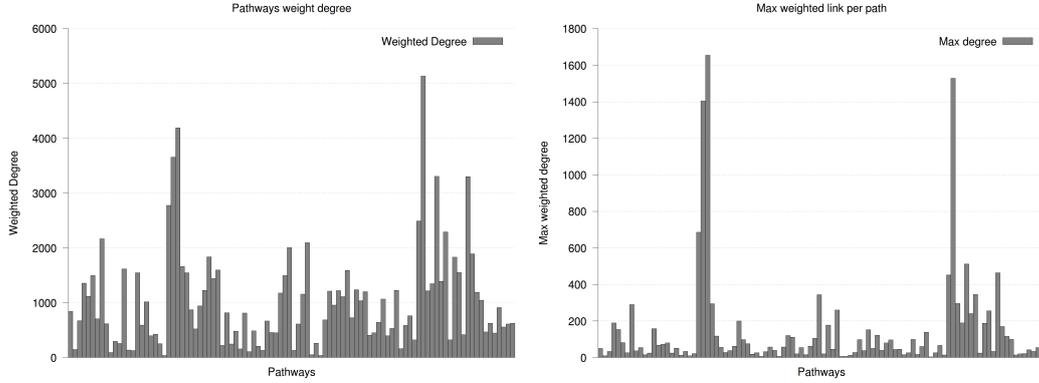


Figura 4.2: Distribuzione dei link pesati nella rete pathway-pathway (a sinistra) e valore del link con più peso per pathway (destra).

7805 links e una densità del 78.05% e connettività media non pesata di 82.5.

## 4.2 Misura alternativa per la forza delle connessioni

Il network così costituito è quindi molto connesso: per questo motivo si introduce una misura aggiuntiva per stabilire quanto due nodi siano legati fra loro. La matrice viene inizialmente simmetrizzata affinché i link siano bi-direzionali. A questo punto ogni connessione fra i pathway  $i$  e  $j$  viene divisa per la somma del numero dei metaboliti unici appartenenti a  $i$  e a  $j$ :

$$L_w^{ij} = \frac{O_w^{ij}}{M_t^{ij}} \quad (4.1)$$

dove  $i$  e  $j$  rappresentano gli indici dei nodi,  $L_w$  rappresenta il link riscalo ( $0 \leq L_w \leq 1$ ),  $O_w$  il numero di metaboliti condivisi secondo la definizione con cui abbiamo creato il network pathway-pathway ed infine  $M_t$  è il numero della somma dei metaboliti posseduti da  $i$  e  $j$ . Sommando  $L_w^{ij}$  su  $j$  otteniamo  $L_w^{tot,i}$  (eq 4.2), che rappresenta la forza di connessione di ogni singolo nodo  $i$ ,

grandezza con la quale possiamo ordinare i pathway.

$$L_w^{tot,i} = \sum_{j=1}^{100} L_w^{ij} \quad (4.2)$$

La figura 4.3, realizzata con il software Gephi [25], rappresenta il network pathway-pathway simmetrizzato con i nodi ed i link pesati secondo la misura che è appena stata introdotta: i link sono di spessore diverso in base a  $L_w^{ij}$ , i nodi sono disposti radialmente in senso orario a decrescere secondo  $L_w^{tot,i}$  e posizionati lungo i rispettivi assi secondo la connettività classica (numero di link) a diminuire dal centro verso l'esterno. Il pathway con valore di  $L^{tot}$

Pathways	Nm	Nr	D	Dw	Ltotw	SS
glutathione metabolism	33	16	92	1227	11.6%	1.2%
citric acid cycle	58	20	91	1545	11%	1.5%
propanoate metabolism	41	13	92	1110	9.5%	1%
glycolysis/gluconeogenesis	84	40	93	1599	9%	1.6%
nad metabolism	50	23	93	1159	9%	1.2%
pyruvate metabolism	54	32	90	1204	9%	1.1%
phenylalanine metabolism	32	10	88	957	8.9%	0.9%
folate metabolism	87	59	93	1549	8.7%	1.5%
tryptophan metabolism	117	70	92	1832	8.4%	1.9%
purine catabolism	91	38	96	1592	8.4%	1.6%

Tabella 4.1: I primi dieci pathways ordinati secondo la misura  $L^{tot,i}$ . Nm: numero di metaboliti, Nr: numero di reazioni, D: connettività, Dw: connettività pesata,  $L^{tot,i}$ : forza di connessione, SS: componente del vettore stato stazionario.

più alto è il metabolismo del glutathione nonostante conti solo 33 metaboliti e 16 reazioni: la misura che abbiamo introdotto ci informa quindi che questo pathway condivide in rapporto il numero più alto di componenti, rendendolo il più legato del network per quanto non sia fra i pathway più corposi. Il glutathione gioca un ruolo cruciale nelle difese antiossidanti, nel metabolismo

dei nutrienti e nella regolazione di eventi cellulari [26] e  $L^{tot}$  sottolinea questa sua rilevanza nell'equilibrio della rete del metabolismo umano.

### 4.3 Analisi stocastica

La rete dei pathways può essere usata per generare dinamiche stocastiche su di essa: il network originale diretto introdotto in 4.1 può essere infatti interpretato come matrice di transizione di un processo stocastico, dove i pesi dei links sono convertiti in probabilità di transizione di una catena di Markov, una volta che si sono normalizzate ad uno le colonne della matrice di adiacenza del network (in questo caso si parla di matrice stocastica sinistra). Una matrice di questo tipo rappresenta una catena di Markov omogena a stati discreti, per la quale esiste un risultato sulla distribuzione di equilibrio e cioè lo stato a cui il sistema tende all'applicazione delle transizioni probabilistiche. Indicando con  $A$  la matrice di transizione di probabilità:

$$\lim_{k \rightarrow +\infty} A^k = \lambda \quad (4.3)$$

dove  $\lambda$  rappresenta lo stato stazionario. Innanzitutto l'esistenza per questa matrice di autovettore unico con autovalore e componenti positive è garantito dal teorema di Perron-Frobenius, il quale afferma:

**Theorem 2.** *Sia  $A$  una matrice non negativa, primitiva ed indecomponibile, allora:*

- *l'autovalore massimo  $\lambda$  di  $A$  è reale e positivo*
- *esso è autovalore semplice*
- *l'autovettore corrispondente ha tutte le componenti positive*
- *l'autovettore corrispondente è l'unico autovettore non negativo*

La matrice che descrive il network rispetta le ipotesi del teorema (i pesi sono positivi ed il network è connesso). Nel caso di una matrice di probabilità il

vettore  $\lambda$  è unitario e le sue componenti rappresentano l'unico stato stazionario del sistema (in tabelle 4.2 e 4.3 riportato sotto SS) coincidendo perciò con il vettore  $\lambda$  dell'eq 4.3. Le componenti di SS possono essere usate come una differente misura per il ranking dei pathways, come mostrato in tabella 4.3. Questo ranking è completamente diverso da quello precedente, sottolineando ruoli funzionali di un pathway all'interno del network. In questo caso

<b>Pathways</b>	<b>Nm</b>	<b>Nr</b>	<b>D</b>	<b>Dw</b>	<b>Ltotw</b>	<b>SS</b>
transport, extracellular	940	1550	94	5134	3.8%	5.2%
exchange/demand reaction	810	742	92	3655	2.8%	4%
fatty acid oxidation	960	869	95	4190	3.1%	3.9%
transport, mitochondrial	366	309	93	3307	6.2%	3.4%
unassigned	317	173	97	3301	7%	3.3%
eicosanoid metabolism	391	257	94	2778	4.9%	2.8%
transport, endoplasmic reticular	294	161	92	2490	5.3%	2.5%
transport, peroxisomal	224	126	90	2292	6.1%	2.3%
nucleotide interconversion	206	177	97	2097	6.3%	2.1%
bile acid synthesis	184	128	94	2165	7%	2.1%

Tabella 4.2: I primi dieci pathways ordinati secondo il vettore dello stato stazionario.

il valore di SS risente maggiormente del numero di metaboliti e reazioni che compongono un pathway, in linea col fatto che le informazioni che si ricavano dallo stazionario riguardano la probabilità, per un processo, di terminare in uno specifico pathway cellulare. Possibili applicazioni per questa misura riguardano la possibilità di perturbare il network (ad esempio rimuovendo una reazione, un metabolita o un gene) e valutare l'impatto di tale modifica nei valori dello stato stazionario il quale fornisce una lettura immediata su quali pathways sono stati maggiormente interessati dalla perturbazione.





## Capitolo 5

# Un'applicazione FBA-FVA: ipossia cellulare

*'The modelling process itself is more important than the model. [...] In light of the complexity of molecular systems and the available experimental data, Systems Biology is the art of making the right assumptions in modelling.'* [31]

La modellizzazione matematica fornisce strumenti indispensabili con cui effettuare ed iterare esperimenti virtuali: l'esplicitazione formale delle ipotesi del modello migliora la comprensione del sistema biologico e può svelare punti critici del sistema non noti precedentemente. L'analisi del modello può rivelare fenomeni biologici nascosti o chiarire meccanismi contro intuitivi. Inoltre, sono da considerare la diminuzione in termine di costi economici e temporali, un'aumentata velocità di esecuzione e la possibilità di rimpiazzare esperimenti pericolosi e non conformi all'etica. In questo capitolo viene esposto il risultato di un esperimento in silico a partire dalla ricostruzione metabolica Recon2, che costituisce il nostro modello: su questo network che rappresenta una cellula virtuale sono state scelte ipotesi di lavoro su base fisiologica, le cui conseguenze ed eventuali implicazioni vengono approfondite.

## 5.1 Metabolismo delle cellule cancerose: effetto Warburg

Il processo a più steps della carcinogenesi è spesso descritto come successione temporale di mutazioni somatiche, in maniera analoga ad un processo Darwiniano, dove caratteristiche fenotipiche sono mantenute, guadagnate o perse a seconda del contributo alla fitness individuale. In accordo con questo modello, tratti invasivi nei tumori devono emergere come meccanismo di adattamento all'ambiente proliferativo durante lo svilupparsi della patologia. D'altra parte, la comparsa diffusa di tratti fenotipici in comune nella popolazione cancerosa suggerisce che questi aspetti condivisi debbano conferire un vantaggio selettivo di sorta. Una proprietà comune dei tumori più invasivi è l'alterazione del metabolismo del glucosio. La glicolisi richiede per prima cosa la conversione di glucosio in piruvato e successivamente la trasformazione in acido lattico, che costituisce prodotto di scarto per la cellula. Nella maggior parte delle cellule dei mammiferi la glicolisi è inibita dalla presenza di ossigeno, consentendo alla respirazione mitocondriale di ossidare il piruvato in  $\text{CO}_2$  e  $\text{H}_2\text{O}$ . Questa inibizione è chiamata Pasteur Effect, dal nome dello scopritore. Questa versatilità metabolica è essenziale per il mantenimento della produzione di energia su tutto il range delle concentrazioni di ossigeno disponibili alla cellula. La conversione di ossigeno in acido lattico anche in presenza di ossigeno è conosciuta come glicolisi anaerobia o Warburg effect: l'incremento di questo processo rispetto valori standard è osservato nelle cellule cancerose ed è considerata una caratteristica basilare della stragrande maggioranza dei tumori. L'analisi di migliaia di pazienti in oncologia con tecniche PET hanno mostrato uptake evidente di glucosio, composto fondamentale nel processo metabolico della glicolisi anaerobia. Diversi studi hanno correlato in maniera consistente l'aggressività dei tumori con l'uptake di glucosio ed inoltre i tumori che si sviluppano in ambiente privo di ossigeno (detti ipossici) sono anche più invasivi e metastatizzati. Studi su linee cellulari coltivate hanno trovato che le cellule tumorali mantengono il

loro metabolismo alterato anche in presenza di ossigeno, suggerendo in definitiva che l'osservazione universale della glicolisi anaerobia anche in cellule cancerose in condizioni normossiche e la correlazione con l'aggressività del tumore indicano che il fenotipo glicotico conferisce un significativo vantaggio proliferativo durante l'evoluzione somatica della cellula del cancro ed è per questo un componente cruciale del fenotipo maligno. Quanto esposto va contro immediate osservazioni: il metabolismo anaerobio del glucosio produce come risultato solo due molecole di ATP, dove un completo ciclo di ossidazione nei mitocondri ne produce 38 (figura 5.1). Inoltre i prodotti metabolici

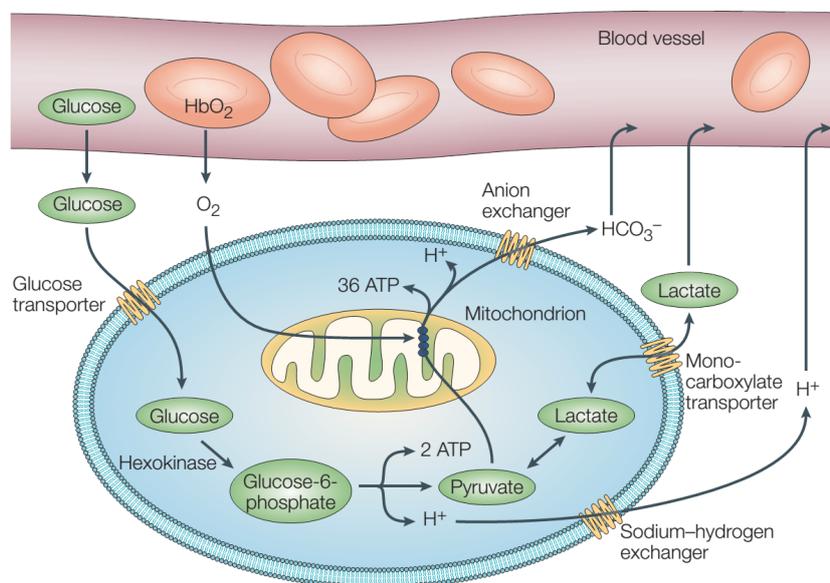


Figura 5.1: Respirazione cellulare: via aerobia e via anaerobia. Da [32].

della glicolisi, come lo ione idrogeno H<sup>+</sup>, causano una consistente acidificazione dello spazio extra-cellulare, aumentandone la tossicità. Per quale motivo quindi le popolazioni tumorali scelgono il fenotipo glicotico, che è inefficiente e potenzialmente tossico? Un'ipotesi al vaglio prevede che la prevalenza del metabolismo up regolato della glicolisi rappresenti una soluzione per superare i vincoli imposti dall'ambiente tumorale stesso e che la sua persistenza in tumori completamente sviluppati indichi il vantaggio proliferativo che conti-

nua a conferire a questo tipo di cellule [32]. In particolare l'ambiente in cui si sviluppa un tumore è generalmente meno vascolarizzato, almeno nelle fasi iniziali, rispetto al tessuto sano: all'insorgenza di una patologia tumorale, i vasi sanguigni restano confinati nello stroma, ossia il tessuto connettivo di supporto ad ogni organo, costringendo le cellule cancerose ad evolversi in condizioni di ipossia e cioè in scarsità di ossigeno (figura 5.2). La successiva proliferazione non farà altro che allontanare ulteriormente le cellule cancerose dai vasi sanguigni. In questo ambiente la selezione microambientale favorirà lo sviluppo di fenotipi capaci di adattarsi alle difficili condizioni in cui si trova il tessuto tumorale, privilegiando quelle cellule capaci di resistere all'ipossia indotta. Successive considerazioni riguardano la pressione selettiva

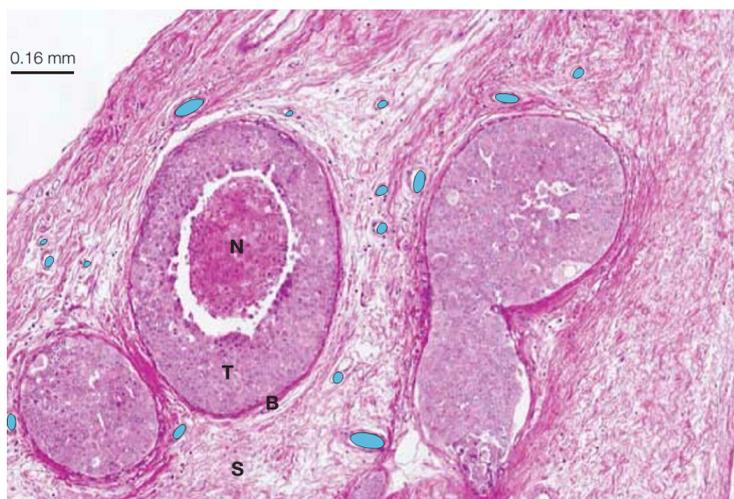


Figura 5.2: I vasi sanguigni in blu sono immersi nello stroma (S) che circonda il tessuto tumorale (B): il tumore non è vascolarizzato, il centro del tumore è necrotico (N). Da [32].

che questo ambiente induce sulla cellula: per proliferare devono essere silenziati tutti quei meccanismi di feedback che nel caso non patologico inducono l'apoptosi in mancanza di ossigeno, come l'hypoxia-inducible factor e all'opposto dovranno essere mantenuti attivi quei geni che regolano la crescita e la duplicazione come HRAS o KRAS. Diversi studi hanno dimostrato che

una bassa concentrazione di ossigeno sia la prima limitazione che si presenta ad una cellula tumorale, suggerendo che in effetti l'ipossia nelle regioni pre-tumorali produce quell'insieme di condizioni che favoriscono il passaggio al metabolismo anaerobico, che consente il mantenimento delle funzioni cellulari anche in carenza/mancanza di ossigeno [33][34][35][36].

## 5.2 L'ipossia nel modello Recon2

L'esperimento in silico che verrà ora esposto si basa sulle considerazioni appena riportate riguardo il metabolismo delle cellule cancerose: se infatti la carenza di ossigeno pare essere la condizione iniziale comune alla maggior parte dei tumori (sicuramente la quasi totalità di quelli solidi) potrebbe essere interessante studiare come si comporta la cellula virtuale modellizzata da Recon2 in assenza di ossigeno esterno. Le ipotesi possono essere riportate sul modello in maniera piuttosto rapida:

1. si identifica il metabolita che corrisponde all'ossigeno nello spazio extra cellulare
2. si modificano a piacere i limiti dei flussi delle reazioni di cui è reagente
3. tramite un'analisi della differenza di flussi rispetto al modello non perturbato (baseline) si stimano le conseguenze della diversa disponibilità d'ossigeno sul modello
4. si ripete la procedura per tutti i valori di ossigeno di cui si vuole interessare

La tabella A.2 riporta le 12 reazioni di Recon2 in cui il metabolita ossigeno extra-cellulare, identificato come  $O_2[e]$ , prende parte: è inoltre segnalato il pathway di appartenenza della reazione e la formula chimica completa della reazione. L'esperimento consiste a questo punto nel variare dal minimo al massimo i flussi disponibili per queste reazioni, quindi ripetere l'esperimento coprendo il range  $[0,1000]$  per valutare come la perturbazione ha impattato sul modello tramite l'analisi dei flussi risultanti sulle altre reazioni.

### 5.3 Cobra toolbox e FBA

Per prima cosa si comincia con un'analisi FBA della baseline, ossia del modello non perturbato di Recon2, importando in Matlab la ricostruzione così come viene fornita. A questo punto si attiva il toolbox Opencobra [27], il quale raccoglie le principali routines dei metodi di analisi di modelli metabolici basate su vincoli. I comandi utilizzati possono essere tutti trovati sulla guida redatta dal gruppo di ricerca che ha sviluppato il toolbox, in [37]. Il pacchetto software sceglie automaticamente il software risolutore di problemi fra quelli disponibili sul computer di linear programming. In questo caso il primo risolutore utilizzato è quello fornito dalla comunità GNU, ossia GLPK [28]. L'analisi FBA richiede un solo comando per essere eseguita e ritorna una struttura che contiene il valore massimizzato della funzione obiettivo (che nel nostro caso è la biomassa, come spiegato in 3.2) e il vettore dei flussi passanti per le singole reazioni. Il risultato trovato numericamente per quanto riguarda il flusso passante per la biomassa, quindi il valore della massimizzazione della funzione obiettivo, risulta essere 3.1981 mmol/g DCW/h (millimoli su grammo per dry cell weight su ora): l'unità di misura di questo valore di flusso è dettata dai coefficienti stechiometrici attribuiti ai metaboliti che compongono la reazione biomassa (che ricordiamo essere stata compilata da dati reali) e risulta essere biologicamente molto alto. Questo risultato è in linea con il fatto che recon2 viene fornito senza nessun particolare vincolo sui flussi interni ed esterni delle reazioni, cosicché nel modello virtuale la cellula è libera di pescare nutrienti e condurre le varie reazioni attingendo da fonti pressoché infinite. Più interessante quindi del valore massimo trovato dalla massimizzazione è il vettore dei flussi: della stessa dimensione del vettore delle reazioni di recon (7440), questo vettore contiene indicazioni su quali reazioni sono state attivate e con quali flussi. Si procede quindi con un test di controllo, mantenendo la versione non perturbata di recon2 ma cambiando il risolutore di problemi lineari. Utilizzando il software Gurobi [29], il valore di flusso trovato al termine della massimizzazione della biomassa è il medesimo di quello precedente, ma un'analisi dei flussi rileva differenze notevoli.

La tabella 5.1 mostra alcuni esempi riguardo la disparità di flussi restituiti dai due risolutori lineari. E' evidente che i due risolutori applicano algorit-

Reazione	Glpk	Gurobi
11DOCRTSLtm	-606.59	0
25HVITD2tm	0	1000
2HCO3_NAt	-812.47	-526.62
3HBCDm	0.16	0
ARACHt	-1000	2

Tabella 5.1: Diffenti valori di flusso (in mmol/g DCW/h) a seguito di massimizzazione FBA fra i risolutori glpk e gurobi.

mi diversi fra loro, il cui risultato FBA è il medesimo (stesso valore per la funzione obiettivo) ma che forniscono risultati completamente diversi e non paragonabili fra loro per il vettore dei flussi. Questa situazione è compatibile con quanto visto già in 2.3 e riportato ad esempio in [10]: in network metabolici di grosse dimensioni è possibile che più di una soluzione porti allo stesso risultato di massimizzazione per la biomassa. Questo succede a causa di pathways ridondanti in una o più funzioni o per mancanza di una definizione più precisa dei boundaries di partenza imposti alle reazioni. Nel nostro caso si tratta quasi sicuramente di una combinazione di questi due fattori. A questo punto si abbandona l'analisi con FBA, che non risulta essere più affidabile per gli scopi che ci siamo preposti, e l'indagine passa ad un'analisi di tipo FVA.

## 5.4 Introduzione della FVA nell'indagine

Esposti i limiti dell'utilizzo della FBA, la FVA risulta essere la soluzione ideale: in effetti questa tecnica altro non è che un'applicazione di FBA in cui vengono ricavati gli estremi superiori ed inferiori di flusso per ogni reazione a parità di ottimo della funzione obiettivo. In sostanza, viene riportato l'inte-

ro cono di soluzioni nello spazio vettoriale in cui la funzione obiettivo viene massimizzata secondo FBA. La versione FVA inclusa nel pacchetto matlab di opencobra risulta essere estremamente lenta per l'obiettivo di questa indagine: infatti, la FVA è sicuramente molto più onerosa a livello computazione di FBA e l'algoritmo proposto in opencobra non è ne ottimizzato ne pronto ad essere parallelizzato nonostante il problema di per sé altro non sia che una profonda parallelizzazione di FBA. A tale scopo viene quindi preferito l'algoritmo fastFVA [38] il quale si può interfacciare ad i più veloci risolutori di problemi lineari e consente la parallelizzazione sul numero di processori disponibili per il calcolo. Viene inoltre fornito in codice c compilato con interfaccia Matlab, in modo da essere eseguito in maniera molto più efficiente di un'eventuale controparte scritta in puro codice Matlab. Il software impiegato per risolvere il problema di programmazione lineare è stavolta fornito dall'IBM e si tratta di Cplex [30], che fra l'altro costituisce lo standard de-facto per la risoluzione di problemi di questo tipo in ambito biologico. Un'ispezione del codice sorgente del programma fastFVA rivela che l'algoritmo contiene un parametro che modifica la tolleranza con il quale l'analisi FVA riconosce come ottimo il risultato della massimizzazione: questo vuol dire che tutti i flussi che portano ad un risultato compreso fra l'ottimo e il range di tolleranza previsto verranno accettati come soluzioni possibili del problema di massimizzazione. Il valore di tolleranza con cui il software viene fornito è  $10^{-6}$ : questa scelta, che può avere senso in un problema in silico ben posto dove tutti i boundaries attribuiti alle reazioni sono stati individuati da misure di laboratorio, risulta essere stringente nel nostro caso. Si decide quindi di eliminare l'ambiguità dovuta all'introduzione di una tolleranza sul risultato, modificando il codice sorgente in maniera tale da accettare un ulteriore parametro con il quale variare la tolleranza: ricompilando successivamente il codice otteniamo una versione adeguata alle nostre esigenze. Per garantire di esserci posti in una condizione definita sia dal punto di vista del calcolo numero che dei risultati vengono eseguiti diversi run a diversa tolleranza, cercando il valore minimo a cui si trova la convergenza dei risultati. In ta-

Tolleranza	Lower Bound	Upper Bound
1 - 1E-6	7049	6395
1E-6 - 1E-10	7249	6851
1E-10 - 1E-15	7255	7215
1E-15 - 1E-30	7440	7320
1E-30 - 1E-60	7440	7440

Tabella 5.2: Convergenza dei risultati FVA alla variazione della tolleranza: dalla soglia di 1E-30 in poi i flussi convergono allo stesso risultato.

Tabella 5.2 vengono riportati i risultati dell'analisi così effettuata: si è variata la tolleranza da 1 a  $10^{-60}$  e per ogni run si sono contati i risultati di flussi per ogni reazione trovati con il medesimo del run precedente. Risulta che per tolleranze alte c'è anche molta variazione nei risultati, cosicché ad esempio fra i run lanciati con tolleranza 1 e  $10^{-6}$  si trovano 7049 flussi uguali per il lower bound e 6395 per l'upper bound FVA. Calando la tolleranza al livello di precisione numerica della macchina si trova che da  $1^{-30}$  in poi i risultati convergono sia per l'upper bound che il lower bound restituiti dal FVA. Si decide quindi di continuare l'analisi ponendo la tolleranza a  $1^{-30}$ .

## 5.5 Simulazione e risultati

La simulazione vera propria si è svolta modificando i flussi per le reazioni con reagente ossigeno extracellulare nel range fra [0,1000] a passi di 0.5: in totale sono stati quindi eseguiti 2000 lanci FVA. Questa scelta è stata fatta per coprire nel dettaglio tutto lo spettro dei valori attribuiti ai flussi delle reazioni Recon. E' da notare che questi valori sono stati assegnati in accordo con i segni dei boundaries inferiori e superiori: quindi i lower boundaries di ogni reazione hanno segno negativo e gli upper boundaries hanno stesso valore ma con segno positivo (esempio: lower bound: -50, upper bound: +50). Il vettore dei flussi per ognuno dei 2000 lanci è stato salvato per le

analisi successive. Prima di tutto è utile chiarire il criterio con cui sono state analizzate le differenze fra i lanci ad ossigeno variabile e la baseline Recon2: si sono considerati separatamente upper bound e lower bound, valutando per ciascuno di essi quando il flusso nella simulazione ad ossigeno variato è aumentato o diminuito rispetto la baseline. In questo modo per ogni lancio si sono ottenuti 2 vettori contenenti le reazioni a flusso maggiore o minore rispetto l'originale Recon2. A questo punto è possibile mostrare i risultati di tale analisi. Il grafico 5.3 mostra il numero di reazioni con flusso aumentato

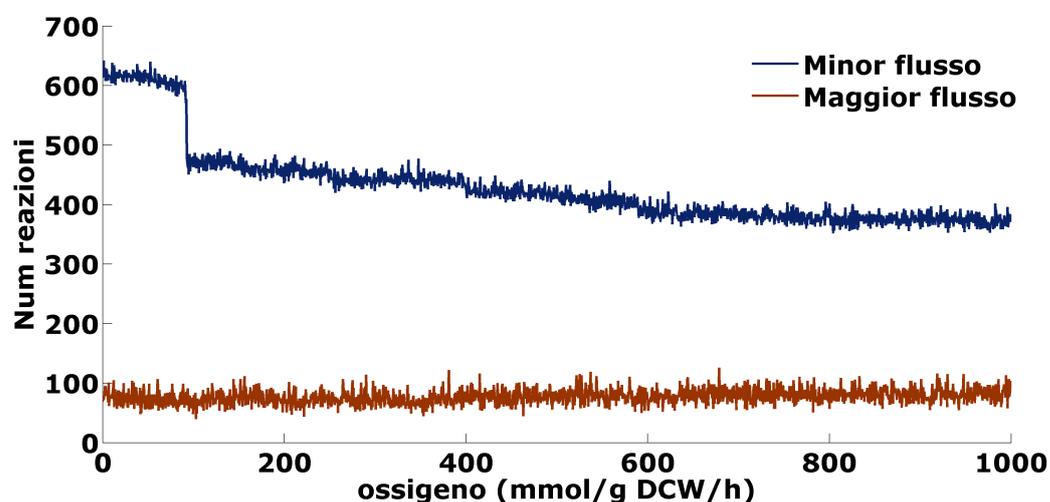


Figura 5.3: Numero di reazioni con flusso diminuito e aumentato per run.

e diminuito per tutti i 2000 lanci di FVA. Notiamo subito che il numero di reazioni a flusso aumentato resta pressoché costante lungo tutti i run e questo può voler dire che la rimozione dell'ossigeno (di qualsiasi entità) influenza in maniera positiva sempre circa lo stesso gruppo di reazioni o molto più probabilmente la perturbazione introdotta non riesce effettivamente ad indurre un aumento dei flussi proporzionato alla modifica e che quindi ci troviamo di fronte ad una variazione introdotta dalla precisione numerica con cui viene processato il problema di FVA, indipendentemente dall'entità della rimozione di ossigeno. A supporto di questa seconda ipotesi è importante notare come la variazione in positivo del valore assoluto dei flussi delle reazioni è

sempre nei pressi della tolleranza impostata dalla FVA, dell'ordine di  $10^{-20}$  -  $10^{-30}$ . Nessuna reazione aumenta il proprio flusso in tutti i 2000 run, mentre se prendiamo come riferimento una variazione in positivo in almeno 1000 run otteniamo che 47 reazioni sono aumentate in termini di flusso. Nella tabella 5.3 sono elencati quei paths che contengono le reazioni a flusso aumentato in almeno 1000 run. Il motivo per cui sia difficile appurare in maniera oggettiva

Pathways
blood group synthesis
cholesterol metabolism
exchange/demand reaction
keratan sulfate degradation
keratan sulfate synthesis
lysine metabolism
n-glycan degradation
n-glycan synthesis
nad metabolism
pentose phosphate pathway
sphingolipid metabolism
transport, extracellular
transport, golgi apparatus
transport, lysosomal
transport, mitochondrial
transport, peroxisomal
tryptophan metabolism
ubiquinone synthesis

Tabella 5.3: Paths con reazioni a flusso aumentato in almeno 1000 run.

quali reazioni e quindi quali processi sono stati interessati da un aumento di flusso sistematico è da ricercarsi ancora una volta nei boundaries di Recon2: come è stato segnalato in 3.1, i vincoli sui flussi del modello originale sono in

gran parte rimossi, ossia la maggioranza delle reazioni ha già a disposizione il massimo della capacità di flusso possibile. Questo andamento costante si riflette anche nei pathways in cui vengono ripartite le reazioni a flusso aumentato, come si può vedere in 5.4. In effetti, sempre dal grafico 5.3, la

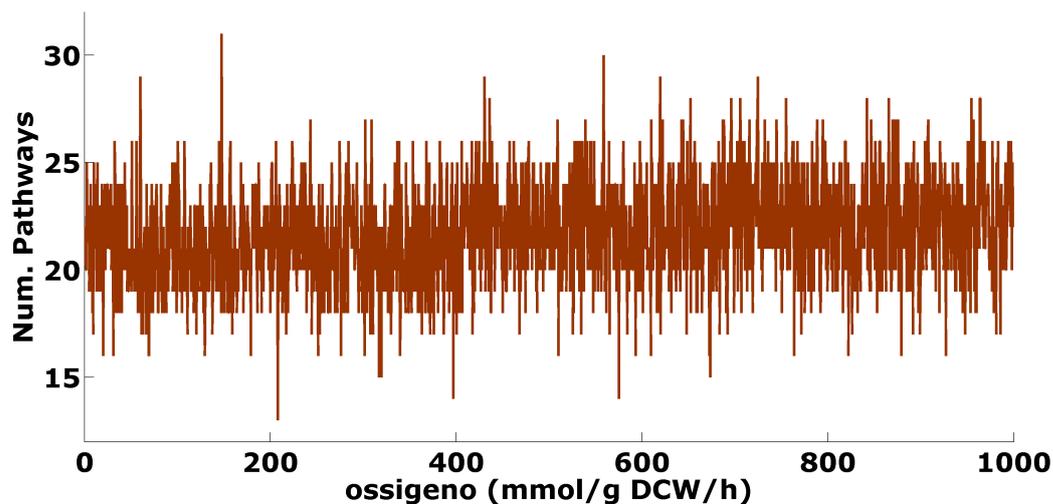


Figura 5.4: Pathways con reazioni a flusso aumentato.

dinamica che più risalta è quella della diminuzione dei flussi. Si nota infatti che a mano a mano che l'ossigeno viene rimosso, il numero di reazioni a flusso diminuito aumenta e nei dintorni di 90 mol/g DCW/h, quindi di un decimo del flusso massimo a disposizione delle reazioni con ossigeno extracellulare, si ha un netto stacco che porta il numero di reazioni a flusso diminuito da poco meno di 500 a poco più di 600. Le reazioni che diminuiscono il proprio flusso in tutti i 2000 run sono 247 e vengono distribuite in pathways come mostrato in tabella 5.4. Analogamente all'andamento appurato in 5.3 è anche il comportamento del numero dei pathways in cui vengono ripartite le reazioni a flusso diminuito lungo i run, come mostrato in figura 5.5. Risulta quindi più interessante valutare l'andamento del flusso totale delle reazioni a flusso diminuito rispetto al totale del flusso originale delle medesime reazioni della baseline Recon2, come viene mostrato nel grafico 5.6. Risulta evidente che se prendiamo in considerazione le variazioni in negativo del flusso, emerge un

<b>Pathways</b>
androgen and estrogen synthesis and metabolism
arginine and proline metabolism
bile acid synthesis
cholesterol metabolism
exchange/demand reaction
fatty acid oxidation
fatty acid synthesis
heme degradation
methionine and cysteine metabolism
sphingolipid metabolism
squalene and cholesterol synthesis
steroid metabolism
taurine and hypotaurine metabolism
transport, endoplasmic reticular
transport, extracellular
transport, golgi apparatus
transport, mitochondrial
transport, nuclear
transport, peroxisomal
tyrosine metabolism
urea cycle
vitamin d metabolism

Tabella 5.4: Pathways con reazioni a flusso diminuito in tutti i 2000 run.

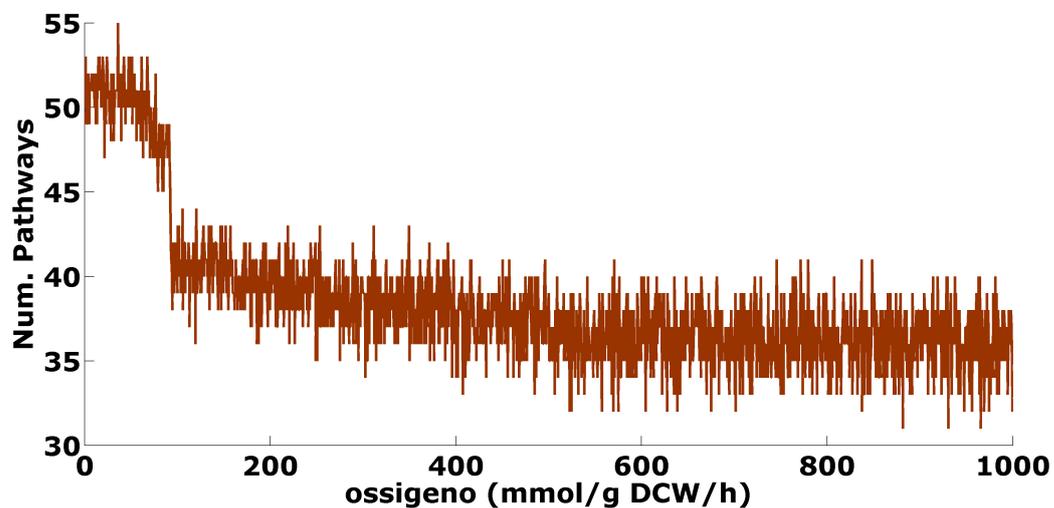


Figura 5.5: Numero pathways con reazioni a flusso diminuito lungo i run.

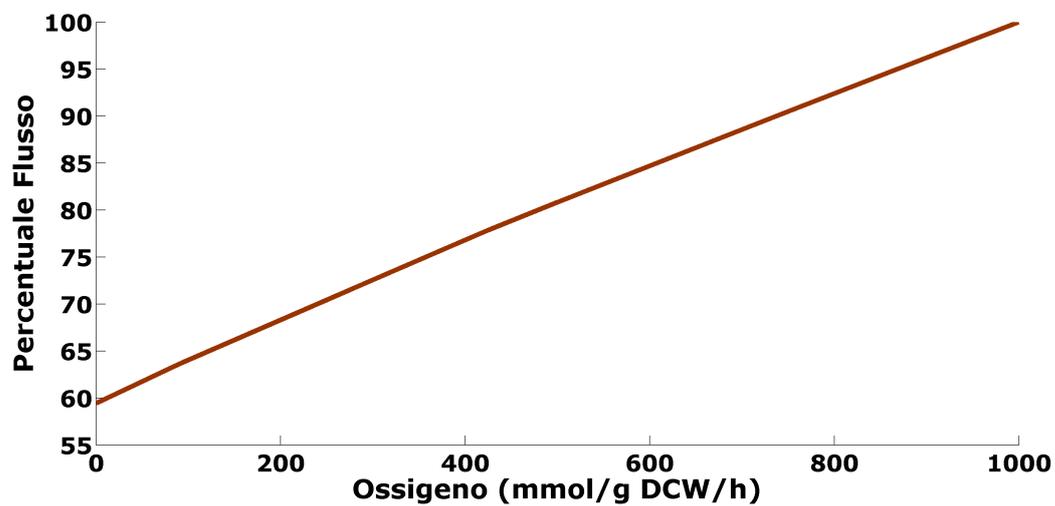


Figura 5.6: Rapporto percentuale fra il flusso totale delle reazioni a flusso diminuito e flusso totale delle stesse reazioni della baseline Recon2.

andamento quasi lineare nel calo: il punto in cui la pendenza del grafico si fa più ripida corrisponde allo stesso in cui nel grafico 5.3 si ha un aumento netto delle reazioni a flusso diminuito, ossia 90 mmol/g DCW/h. Un ultimo risultato interessante sempre in termini di reazioni a flusso diminuito riguarda la visualizzazione totale del numero percentuale di reazioni a flusso calato per pathway rispetto il numero totale di reazioni di quel pathway per tutti i run, come mostrato in figura 5.7. In 5.7 si notano alcune cose:

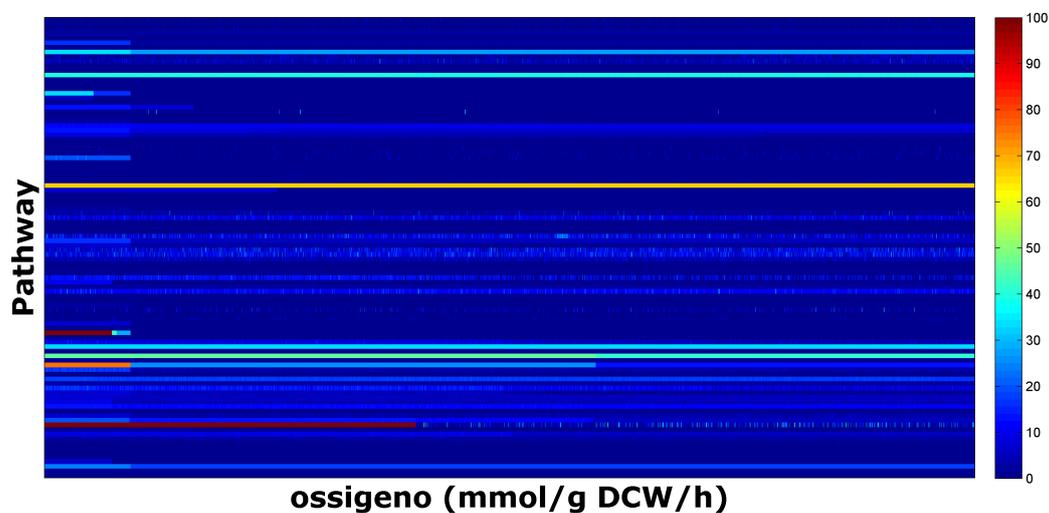


Figura 5.7: Percentuale di reazioni a flusso calato rispetto il numero totale di reazioni per pathways, lungo tutti i run.

- 30 pathways (elencati in A.3) non vengono interessati dal calo dell'ossigeno disponibile
- alcuni pathways si mantengono costanti nel numero di reazioni a flusso diminuito lungo tutti i run
- altri pathways, la maggior parte, vengono interessati ad intermittenza
- da un certo valore di ossigeno in poi, due pathways presentano tutte le proprie reazioni in calo di flusso: si tratta dei pathways dell' ubiquinone synthesis e del nucleotide sugar metabolism.



# Conclusioni

La ricostruzione metabolica Recon2 rappresenta lo stato dell'arte della modellizzazione del metabolismo umano in silico. Sebbene tutt'ora incompleta, le informazioni di cui è composta (matrice stechiometrica, annotazioni semantiche, ontologia, geni codificanti, etc...) rendono questo oggetto un valido strumento per simulazioni al computer ed analisi delle relazioni esistenti fra i composti e processi del metabolismo cellulare. Questo lavoro di tesi si pone nello stadio intermedio fra la creazione della ricostruzione e la sua applicazione pratica: si è infatti approfondito lo studio della struttura, il significato dei legami esistenti e gli ambiti di applicazione sia classici pertinenti alla biologia dei sistemi sia quelli più originali della fisica delle reti. In prima battuta è stato necessario capire come il modello è stato costruito ed assemblato nell'insieme: il ruolo della matrice stechiometrica, il significato dei coefficienti, la nomenclatura ed i riferimenti per le reazioni ed i metaboliti. Successivamente si è passati alla comprensione della struttura matematica e dei principi della flux balance analysis, attualmente lo strumento più noto di indagine applicato a questo tipo di strutture. In quest'ottica sono state realizzate simulazioni di FBA sulla baseline Recon2 con software risolutivi diversi, al fine di ottenere una caratterizzazione numerica di questi metodi. Si è appurato che mentre il risultato della massimizzazione della funzione obiettivo (la biomassa) non cambia variando gli algoritmi di calcolo, la soluzione non è univoca ma viene descritta da un cono nello spazio dei vettori di flusso. Tale situazione si verifica in virtù delle dimensioni della matrice stechiometrica, dei vincoli ai flussi che vengono attribuiti di default alla ri-

costruzione (i quali non sono particolarmente stringenti) e della ridondanza dovuta alla presenza di più processi metabolici che completano il medesimo obiettivo (la produzione di uno più metaboliti specifici). Si nota che la FBA resta comunque un metodo valido di analisi dei flussi nel caso vengano rilassate le condizioni appena descritte: su di un network ridotto, senza ridondanza e con boundaries ben attribuiti a tutte le reazioni di scambio, un'analisi FBA ritorna sia la massimizzazione della funzione obiettivo sia un vettore univoco di flussi nello spazio del risultato. Per sopperire alla carenza di specificità trovata dalla FBA si è quindi passati all'introduzione della flux variability, che consente l'individuazione del cono completo di soluzioni nello spazio dei flussi. Anche in questo caso abbiamo preferito cominciare con una caratterizzazione numerica del metodo, vagliandone il codice sorgente. Ci si è quindi resi conto che, così come fornito, il metodo FVA funziona nell'intorno del massimo della funzione obiettivo, avendo introdotto un parametro di tolleranza. Variando questo parametro si è trovato che i risultati delle simulazioni variano di conseguenza: è stato quindi necessario modificare il codice sorgente in maniera da accettare un valore di tolleranza sufficientemente basso da stabilizzare i risultati. Eliminato questo elemento di arbitrarietà è stato finalmente possibile testare il metodo sulla ricostruzione Recon2, con un esperimento in silico suggerito da considerazioni fisiologiche: si è scelto di simulare la condizione iniziale in cui la maggior parte dei tumori si sviluppa, che è quella della mancanza di ossigeno o ipossia. Questa condizione è facilmente modellabile su recon, intervenendo sui flussi delle reazioni che presentano come reagente ossigeno extracellulare. La carenza di ossigeno è stata simulata lungo tutto lo spettro dei valori ammessi da recon, cosicché si è ottenuta un'analisi particolarmente vasta. In tal senso si è trovato che quasi 250 reazioni della ricostruzione presentano valori di flusso diminuiti in tutti i run che compongono la simulazione, mentre per quanto riguarda gli aumenti di flusso non sono stati individuati risultati particolarmente informativi. In ogni caso, si è deciso di proporre una modalità di analisi diversa da quella della singola reazione: sfruttando la classificazione con cui recon

etichetta ogni reazione con il suo pathway di appartenenza è possibile spostarsi ad un livello più alto di indagine, riducendo lo spazio dell'analisi ai soli 100 pathways che compongono il modello metabolico. In tal modo sono stati trovati quei 22 pathways (tabella 5.4) che contengono le 250 reazioni di cui sopra: si è ridotto quindi di un fattore dieci il numero di elementi da considerare nel risultato. Mantenendo questo criterio di analisi, è anche possibile graficare l'andamento percentuale del numero di reazioni in diminuzione di flusso per pathway come mostrato in figura 5.7: il risultato più interessante di questa modalità riguarda i pathways ubiquinone synthesis ed nucleotide sugar metabolism, per i quali il 100% delle proprie reazioni cala di flusso all'aumentare dell'ipossia. L'ubiquinone synthesis è quel pathway preposto alla produzione della vitamina Q e l'insieme di processi di cui è composto si svolge nei mitocondri: questi sono anche la sede principale della respirazione aerobia della cellula. Non è quindi da escludersi che il processo di produzione della vitamina q risenta della mancanza di ossigeno per questo motivo. Per quanto riguarda il pathway nucleotide sugar metabolism, esistono paper in letteratura che legano l'ipossia alla modulazione di quello specifico processo [39] [40]. A lato dell'analisi dei flussi con i metodi appena esposti, si è portato avanti parallelamente un'analisi di Recon2 più incentrata sulla fisica delle reti. L'estrazione dei network metaboliti-metaboliti e reazioni-reazioni costituisce un approccio fin'ora originale all'esplorazione delle relazioni esistenti fra i componenti della ricostruzione. Su questi network è stato possibile eseguire alcune basilari misure di centralità in grado di evidenziare i principali metaboliti e reazioni che compongono la rete ed in ogni caso altri tipi di indagini possono essere portate avanti sempre nello schema della fisica dei networks e dei suoi strumenti. In ultimo, sempre ricorrendo alla struttura a pathways che la ricostruzione di Recon2 fornisce indirettamente, è stato possibile estrarre un'ulteriore rete e cioè quella dei pathways stessi: questo network ci informa di quanto un processo sia legato ad un altro in termini di metaboliti prodotti e reagenti. Una descrizione a così alto livello ha il vantaggio di condensare in una struttura 100x100 molta della dinamica dell'intera

ricostruzione metabolica, senza considerare la possibilità di poter graficare con successo il network per intero evidenziando le relazioni e le metriche che abbiamo scelto. E' stata introdotta inoltre una misura originale su questo network, che sintetizza la forza di connessione di ciascun nodo (quindi di ogni pathways) con i suoi vicini e più in generale informa di quanto un nodo è attivo all'interno della rete. Per finire, la matrice d'adiacenza della rete dei pathways è stata interpretata (dopo le dovute operazioni matematiche) come una matrice di transizione di un processo markoviano e i pesi che la compongono sono quindi diventati probabilità. In questo nuovo schema è stato possibile trovare lo stato stazionario del processo caratteristico del network non perturbato di partenza: si suggerisce quindi che questa operazione possa essere eseguita su di un eventuale network dei pathway perturbato (dalla rimozione di un gene, di un metabolita o di una reazione) per poi confrontare lo stato stazionario originale con il nuovo che si ottiene a causa della modificazione introdotta. Questa tipologia di analisi può costituire quindi un metodo di tipo puramente stocastico e indipendente dai flussi per stabilire quanto una perturbazione impatta su Recon2 sulla base del solo network dei pathways.

# Appendice A

## Tabelle di supporto

Tabella A.1: Pathways di recon, con numero reazioni e metaboliti unici

<b>Pathway</b>	<b>Rxns</b>	<b>Mets</b>
Alanine and aspartate metabolism	16	50
Alkaloid synthesis	5	18
Aminosugar metabolism	31	60
Androgen and estrogen synthesis and metabolism	57	119
Arachidonic acid metabolism	73	126
Arginine and Proline Metabolism	41	89
Bile acid synthesis	128	184
Biotin metabolism	12	32
Blood group synthesis	46	56
Butanoate metabolism	3	11
C5-branched dibasic acid metabolism	8	14
Cholesterol metabolism	62	117
Chondroitin sulfate degradation	44	45
Chondroitin synthesis	45	58
Citric acid cycle	20	58
CoA catabolism	6	24

CoA synthesis	20	45
Cysteine Metabolism	3	10
Cytochrome metabolism	15	42
D-alanine metabolism	3	10
Dietary fiber binding	12	21
Eicosanoid metabolism	257	391
Exchange/demand reaction	742	810
Fatty acid oxidation	869	960
Fatty acid synthesis	126	190
Folate metabolism	59	87
Fructose and mannose metabolism	27	42
Galactose metabolism	12	32
Glutamate metabolism	15	38
Glutathione metabolism	16	33
Glycerophospholipid metabolism	74	153
Glycine, serine, alanine and threonine metabolism	37	91
Glycolysis/gluconeogenesis	40	84
Glycosphingolipid metabolism	14	46
Glyoxylate and dicarboxylate metabolism	15	40
Heme degradation	3	14
Heme synthesis	14	29
Heparan sulfate degradation	27	38
Histidine metabolism	16	39
Hyaluronan metabolism	5	12
Inositol phosphate metabolism	64	65
Keratan sulfate degradation	75	76
Keratan sulfate synthesis	59	70
Limonene and pinene degradation	9	27
Linoleate metabolism	24	46
Lipoate metabolism	2	8

---

Lysine metabolism	25	63
Methionine and cysteine metabolism	37	94
Miscellaneous	88	229
N-glycan degradation	16	33
N-glycan synthesis	108	126
NAD metabolism	23	50
Nucleotide interconversion	177	206
Nucleotide salvage pathway	3	8
Nucleotide sugar metabolism	5	15
O-glycan synthesis	15	23
Oxidative phosphorylation	10	24
Pentose phosphate pathway	39	60
Phenylalanine metabolism	10	32
Phosphatidylinositol phosphate metabolism	63	98
Propanoate metabolism	13	41
Purine catabolism	38	91
Purine synthesis	13	32
Pyrimidine catabolism	35	70
Pyrimidine synthesis	19	60
Pyruvate metabolism	32	54
R group synthesis	50	53
ROS detoxification	7	22
Selenoamino acid metabolism	16	50
Sphingolipid metabolism	83	149
Squalene and cholesterol synthesis	6	19
Starch and sucrose metabolism	33	52
Steroid metabolism	76	121
Stilbene, coumarine and lignin synthesis	2	8
Taurine and hypotaurine metabolism	8	15
Tetrahydrobiopterin metabolism	27	43

---

---

Thiamine metabolism	6	15
Transport, endoplasmic reticular	161	294
Transport, extracellular	1550	940
Transport, golgi apparatus	78	160
Transport, lysosomal	107	209
Transport, mitochondrial	309	366
Transport, nuclear	65	130
Transport, peroxisomal	126	224
Triacylglycerol synthesis	13	35
Tryptophan metabolism	70	117
Tyrosine metabolism	121	196
Ubiquinone synthesis	14	30
Unassigned	173	317
Urea cycle	69	135
Valine, leucine, and isoleucine metabolism	42	71
Vitamin A metabolism	81	108
Vitamin B12 metabolism	3	14
Vitamin B2 metabolism	7	21
Vitamin B6 metabolism	11	15
Vitamin C metabolism	16	39
Vitamin D metabolism	29	50
Vitamin E metabolism	24	50
Xenobiotics metabolism	25	52
beta-Alanine metabolism	12	37

Tabella A.2: Reazioni di Recon2 interessate dall'ipossia

	<b>Nome reazione</b>	<b>Pathway</b>
1	exchange reaction for oxygen	Exchange/demand reaction
2	o2 transport (diffusion)	Transport, extracellular
3	Fe(II):oxygen oxidoreductase Porphyrin and chlorophyll metabolism	Miscellaneous
4	Putrescine:oxygen oxidoreductase (deaminating) Urea cycle and metabolism of amino groups	Methionine and cysteine metabolism
5	RE0688	Urea cycle
6	RE0689	Urea cycle
7	RE0690	Urea cycle
8	RE0827	Urea cycle
9	RE0828	Urea cycle
10	RE3367	Urea cycle
11	CYSAMOe	Transport, extracellular
12	HYPTROXe	Methionine and cysteine metabolism
	<b>Formula reazione</b>	
1	$\leftrightarrow \text{o2}[\text{e}]$	
2	$\text{o2}[\text{e}] \leftrightarrow \text{o2}[\text{c}]$	
3	$4 \text{ h}[\text{e}] + 4 \text{ fe2}[\text{e}] + \text{o2}[\text{e}] \leftrightarrow 2 \text{ h2o}[\text{e}] + 4 \text{ fe3}[\text{e}]$	
4	$\text{h2o}[\text{e}] + \text{o2}[\text{e}] + \text{ptrc}[\text{e}] \leftrightarrow \text{nh4}[\text{e}] + \text{h2o2}[\text{e}] + 4\text{abutn}[\text{e}]$	
5	$\text{h2o}[\text{e}] + \text{sprm}[\text{e}] + \text{o2}[\text{e}] \rightarrow \text{nh4}[\text{e}] + \text{h2o2}[\text{e}] + \text{CE1935}[\text{e}]$	
6	$\text{h2o}[\text{e}] + \text{o2}[\text{e}] + \text{spnd}[\text{e}] \rightarrow \text{nh4}[\text{e}] + \text{h2o2}[\text{e}] + \text{CE1940}[\text{e}]$	
7	$\text{h2o}[\text{e}] + \text{o2}[\text{e}] + \text{CE1940}[\text{e}] \rightarrow \text{nh4}[\text{e}] + \text{h2o2}[\text{e}] + \text{CE1943}[\text{e}]$	
8	$\text{h2o}[\text{e}] + \text{o2}[\text{e}] + \text{CE1935}[\text{e}] \rightarrow \text{nh4}[\text{e}] + \text{h2o2}[\text{e}] + \text{CE1936}[\text{e}]$	
9	$\text{h2o}[\text{e}] + \text{o2}[\text{e}] + \text{spnd}[\text{e}] \rightarrow \text{nh4}[\text{e}] + \text{h2o2}[\text{e}] + \text{CE1939}[\text{e}]$	
10	$\text{h2o}[\text{e}] + \text{o2}[\text{e}] + \text{CE1939}[\text{e}] \rightarrow \text{nh4}[\text{e}] + \text{h2o2}[\text{e}] + \text{CE1943}[\text{e}]$	
11	$\text{o2}[\text{e}] + \text{cysam}[\text{e}] \rightarrow \text{h}[\text{e}] + \text{hyptaur}[\text{e}]$	
12	$\text{o2}[\text{e}] + 2 \text{ hyptaur}[\text{e}] \rightarrow 2 \text{ taur}[\text{e}]$	

<b>Pathways</b>
alkaloid synthesis
arachidonic acid metabolism
beta-alanine metabolism
butanoate metabolism
c5-branched dibasic acid metabolism
chondroitin sulfate degradation
chondroitin synthesis
citric acid cycle
cysteine metabolism
dietary fiber binding
glycine, serine, alanine and threonine metabolism
heparan sulfate degradation
histidine metabolism
hyaluronan metabolism
limonene and pinene degradation
linoleate metabolism
lipoate metabolism
nucleotide salvage pathway
nucleotide sugar metabolism
pyrimidine catabolism
r group synthesis
selenoamino acid metabolism
stilbene, coumarine and lignin synthesis
thiamine metabolism
triacylglycerol synthesis
vitamin b12 metabolism
vitamin b2 metabolism
vitamin b6 metabolism
vitamin e metabolism
xenobiotics metabolism

Tabella A.3: Reazioni di Recon2 non interessate dall'ipossia.

# Bibliografia

- [1] Smith E, Morowitz H (2004): ‘Universality in intermediary metabolism’. Proc Natl Acad Sci USA 101 (36): 13168–73.
- [2] G. Facchetti (2013): ‘Computational approaches to complex biological networks’ - Ph.D. Thesis - SISSA.
- [3] Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, et al. (2007): ‘Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox.’ Nature protocols 2: 727-738.
- [4] Palsson BO (2006): ‘Systems Biology - Properties of Reconstructed Networks.’ - Cambridge University Press - ISBN 9780521859035.
- [5] Price, Nathan; Jennifer Reed; Jason Papin; Sharon Wiback; Bernhard O Palsson (2003). ‘Network-based analysis of metabolic regulation in the human red blood cell’. Journal of Theoretical Biology 225 (2): 185-194.
- [6] Stelling, J; Klamt S; Bettenbrock K; Schuster S; Gilles ED (2002). ‘Metabolic network structure determines key aspects of functionality and regulation.’. Nature 420 (6912): 190-193.
- [7] Larhlimi, A.; Bockmayr, A. (2009). ‘A new constraint-based description of the steady-state flux cone of metabolic networks’. Discrete Applied Mathematics 157 (10): 2257-2266.
- [8] Uri Alon. (2007): ‘An Introduction to Systems Biology: Design Principles of Biological Circuits.’, Chapman and Hall - CRC.

- 
- [9] D. Segré, D. Vitkup, and G.M. Church (2002): ‘Analysis of optimality in natural and perturbed metabolic networks.’ *Proc Natl Acad Sci USA*:15112-15117.
- [10] Jeffrey D Orth, Ines Thiele and Bernhard O Palsson (2010): ‘What is flux balance analysis?’ *Nature Biotechnology* 28, 245-248.
- [11] Ranganathan, Sridhar, Patrick F. Suthers, and Costas D. Maranas (2010): ‘OptForce: An Optimization Procedure for Identifying All Genetic Manipulations Leading to Targeted Overproductions.’ *PLoS Comput Biol* 6, no. 4.
- [12] Hong Yang, Charles M. Roth, Marianthi G. Ierapetritou (2009): ‘A rational design approach for amino acid supplementation in hepatocyte culture.’ *Biotechnology and Bioengineering*, 103: 1176-1191.
- [13] Raghunathan, Anu, Sookil Shin, and Simon Daefer (2010): ‘Systems Approach to Investigating Host-pathogen Interactions in Infections with the Biothreat Agent *Francisella*. Constraints-based Model of *Francisella Tularensis*.’ *BMC Systems Biology* 4, no. 1.
- [14] Mahadevan R, Schilling C (2003): ‘The effects of alternate optimal solutions in constraint-based genome-scale metabolic models.’ *Metabolic engineering*, 5(4):264-276.
- [15] Reed J, Palsson B (2004): ‘Genome-scale in silico models of *E. coli* have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states’. *Genome Research*, 14(9).
- [16] Thiele I, Fleming R, Bordbar A, Schellenberger J, Palsson BO (2010): ‘Functional characterization of alternate optimal solutions of *Escherichia coli*’s transcriptional and translational machinery.’ *Biophysical Journal*, in press. PubMed Abstract.
- [17] Kanehisa, M. (2000): ‘Post-genome Informatics’, Oxford University Press.

- [18] Thiele et al., (2013): ‘A community-driven global reconstruction of human metabolism., Nat Biotechnol.’. doi: 10.1038/nbt.2488.
- [19] Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO (2007): ‘Global reconstruction of the human metabolic network based on genomic and bibliomic data.’ Proc Natl Acad Sci USA, 104:1777-1782.
- [20] MATLAB and Statistics Toolbox Release, The MathWorks, Inc., Natick, Massachusetts, United States.
- [21] Maglott D, Ostell J, Pruitt KD, Tatusova T.: ‘Entrez Gene: gene-centered information at NCBI.’ Nucleic Acids Res. 2007 Jan;35(Database issue):D26-31. Epub 2006 Dec 5. PubMed PubMed Central.
- [22] Webb EC (1992). ‘Enzyme nomenclature 1992: recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes.’ San Diego: Published for the International Union of Biochemistry and Molecular Biology by Academic Press.
- [23] Adam M. Feista and Bernhard O. Palsson (2010):‘The Biomass Objective Function’Curr Opin Microbiol.; 13(3):344-349
- [24] Thiele, I., Palsson, B.O. (2010):‘A protocol for generating a high-quality genome-scale metabolic reconstruction.’ Nature protocols 5, 93-121.
- [25] Bastian M., Heymann S., Jacomy M. (2009): ‘Gephi: an open source software for exploring and manipulating networks.’ International AAAI Conference on Weblogs and Social Media.
- [26] Wu, Guoyao; Fang, Yun-Zhong; Yang, Sheng; Lupton, Joanne R.; Turner, Nancy D. (2004): ‘Glutathione metabolism and its implications for health’. The Journal of nutrition 134 (3): 489-92.

- [27] Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, Kang J, Hyduke DR, Palsson BO. (2011): 'Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0.' *Nature Protocols* 6:1290-1307.
- [28] GNU Linear Programming Kit, <http://www.gnu.org/software/glpk/>
- [29] Gurobi Optimization, Inc (2015): 'Gurobi Optimizer Reference Manual.' <http://www.gurobi.com>
- [30] IBM ILOG CPLEX Optimizer: <http://www.ibm.com/>
- [31] O. Wolkenhauer, U. Klingmuller (2004): 'Systems Biology: From a Buzzword to a Life Sciences Approach.' *BIOforum Europe* 4:22-23.
- [32] R.A. Gatenby, R.J. Gillies (2004): 'Why do cancers have high aerobic glycolysis?' *Nature Reviews Cancer* 4 (11), 891-899.
- [33] Griffiths, JR, McIntyre DJ, Howe FA, Stubbs M (2001): 'Why are cancers acidic? A carrier mediated diffusion model for H<sup>+</sup> transport in the interstitial fluid.' *Novartis Found. Symp.* 240, 46-26.
- [34] Helmlinger G, Yuan F, Dellian M, Jain RK (1997): 'Interstitial pH and pO<sub>2</sub> gradients in solid tumors in vivo: high-resolution measurements reveal a lack of correlation.' *Nature med.* 3, 177-182.
- [35] Cheesand TJ, Gillies RJ, Dale EB (1988): 'Optimum fiber spacing in hollow fiber bioreactor.' *Biotechnol. Bioeng.* 32, 938-992
- [36] Secomb TW et al (1998): 'Theoretical simulation of oxygen transport to tumors by three dimensional networks of microvessels.' *Adv. Exp. Med. Biol.* 454, 629-634.
- [37] SA Becker, AM Feist, ML Mo, G Hannum, BO Palsson, MJ Herrgard (2007): 'Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox.' *Nature protocols* 2 (3), 727-738.

- 
- [38] S. Gudmundsson and I. Thiele. (2010): 'Computationally efficient flux variability analysis.' *BMC Bioinformatics*, 11:489.
- [39] K. Shirato, K. Nakajima, H. Korekane, S. Takamatsu, C. Gao, T. Angata, K. Ohtsubo, N. Taniguchi (2011): 'Hypoxic regulation of glycosylation via the N-acetylglucosamine cycle.' *J Clin Biochem Nutr* Jan 29;48(1):20-5.
- [40] P.Kucharzewska, H. C. Christianson, M.Belting (2015): 'Global Profiling of Metabolic Adaptation to Hypoxic Stress in Human Glioblastoma Cells.' *PLoS One*. Jan 29;10(1):e0116740.



# Ringraziamenti

I primi ringraziamenti vanno sicuramente al Prof. Remondini, con il quale nel corso dei mesi si è instaurato un rapporto di collaborazione non comune per un 'semplice' studente. Per la fiducia che mi è stata accordata, lo sprono e gli aiuti che mi sono arrivati gliene sono e gliene sarò sempre enormemente grato. Ringrazio anche il Prof. Castellani per essersi interessato attivamente al mio lavoro, parallelamente al mio relatore. Una menzione di gratitudine va anche alla coordinatrice del corso di Fisica, la Prof.ssa Ercolessi, per aver accolto le istanze riguardo il mio caso.

Ringrazio la mia famiglia, per essermi stata vicina ed avermi appoggiato in tutti i modi anche in questa seconda fase. Spero possano vedere rapidamente i frutti dei loro sacrifici, dopo aver aspettato tanto.

Ringrazio Caterina, per il ruolo speciale che ha nella mia vita e per aver reso sereni i miei momenti d'abbattimento.

Ringrazio Elisa e Martino, con i quali ho condiviso il corso di studi e i mesi di vita vissuta a Bologna. Spero le nostre strade si incrocino di nuovo: nel mio percorso siete stati fondamentali, entrambi col proprio speciale contributo. Vi auguro il meglio.

Ringrazio i miei amici, tutti: siete lo specchio di quello che sono e mi completate nel resto. Non potrei chiedere di meglio.

Infine, dedico questo lavoro ai miei cari che non sono più qui ma che sono sicuro sarebbero felici almeno quanto lo sono io. E lo dedico anche ai nuovi arrivati, che possano conoscere negli anni una versione di me realizzata.