

Macchine molecolari

Gabriele Filosofi

15 Maggio 2003

Abstract

L'obiettivo di questo lavoro è presentare un quadro critico della teoria delle macchine molecolari, così come sviluppata da Thomas D.Schneider in una serie di suoi articoli.

1 Introduzione

All'interno di una cellula singole macromolecole eseguono una varietà di compiti, dalla replicazione al controllo e alla traduzione del materiale genetico, dal monitoraggio delle variabili fisiche, chimiche e biologiche sia intracellulari che extracellulari, alla eliminazione degli agenti patogeni. In un certo senso ciascuna macromolecola deve "prendere delle decisioni", dissipando una certa energia in presenza di rumore.

Come le macchine macroscopiche, anche le macchine molecolari hanno un numero finito di stati interni. Il massimo numero di decisioni binarie che la macchina può prendere è funzione dell'energia P_y che dissipa nel mezzo circostante, dell'energia del rumore termico N_y e del numero di parti in moto indipendente coinvolte nell'operazione, d_{space} :

$$C_y = d_{space} \log_2 \left(1 + \frac{P_y}{N_y} \right) \quad [bit/operazione]$$

Questa *capacità di macchina* è strettamente legata alla *capacità di Canale* di un sistema di telecomunicazione, così come definita da Shannon [Shannon, 1948; Shannon e Weaver, 1949; Shannon, 1949; Pierce, 1980]. Il teorema di Shannon afferma che che:

1. un segnale, con banda in frequenza B , in [Hz], non può trasportare informazione a una velocità R , in bit/s, superiore alla capacità del canale, $C \equiv B \log_2(1 + \frac{P}{N})$: il rumore distrugge la parte in eccesso $R - C$.
2. se $R \leq C$ allora è possibile trasmettere informazioni con un tasso di errore piccolo a piacere, pur di usare una codifica adeguata.

In altre parole, *mentre la capacità del canale è limitata, l'accuratezza della comunicazione può essere illimitata!*.

Applicato al presente contesto, il teorema dice che se la quantità di informazione che la macchina molecolare acquista è $R \leq C_y$, allora la frequenza di errore può essere resa arbitrariamente piccola, purché la macchina stessa sia abbastanza sofisticata. Si può allora capire come possa l'enzima di restrizione *EcoRI*, con estrema precisione, tagliare il DNA solamente in corrispondenza della sequenza 5'GAATTC3', tra le 4096 sequenze possibili di pari lunghezza [Polisky e altri, 1975; Woodhead e altri, 1981; Pingoud 1985].

Ma i concetti di *trasmettitore*, *canale*, *segnale* e *SNR* usati da Shannon non sono immediatamente traducibili in quello che una macchina molecolare come *EcoRI* ha o fa. La corrispondenza però esiste tra una macchina molecolare e il *ricevitore* di Shannon: entrambi acquistano e dissipano energia, effettuano transizioni verso determinati sottostati in presenza di rumore [Johnson, 1987] e con un tasso finito di errore. *Mentre nella teoria della comunicazione i bit misurano la possibilità di scelta tra simboli, nella teoria delle macchine molecolari i bit misurano la possibilità di scelta tra stati.*

Dal punto di vista del linguaggio scientifico, il paradigma usuale per descrivere la corrispondenza stereospecifica tra biomolecole, per esempio un enzima e il suo substrato, è quello della corrispondenza chiave-serratura [Rastetter, 1983; Gilbert e Greenberg, 1984]. Se la macchina molecolare è vista come un insieme di parti indipendenti in moto Browniano (nel caso più semplice assimilabili a oscillatori classici) le variabili numeriche che descrivono la sua configurazione istantanea possono assumere un numero finito di valori all'interno di una sfera di raggio $\sqrt{N_y}$ nello spazio di velocità. Lo stesso avviene in meccanica statistica, per la distribuzione di velocità delle particelle di un gas.

Quando la macchina molecolare è portata in uno stato "eccitato" acquista energia e la sfera si espande. Quando la macchina esegue la sua operazione specifica, dissipa energia e la sfera si contrae, mentre il suo centro si sposta in una nuova posizione (un nuovo stato). Il numero massimo di transizioni differenti che la macchina può effettuare, a partire dallo stato eccitato, dipende da quante delle sfere più piccole possono essere contenute nella sfera più grande senza intersecarsi. Il logaritmo di questo numero è C_y .

Il modo in cui la selezione naturale ha disposto le sfere più piccole di una macchina molecolare corrisponderebbe alla progettazione del codificatore in un sistema di comunicazione [Sloane, 1984; Cipra, 1990].

Il modello geometrico appena delineato è lo stesso utilizzato da Shannon.

2 Definizione ed esempi di MM

2.1 Definizioni

Definizione 2.1.1. *Una macchina molecolare (Molecular Machine, **MM**) è una singola macromolecola o un complesso macromolecolare [Schneider, 191]*

Oss. 2.1.2. Qui noi non ci interessiamo agli effetti dell'azione congiunta di un gran numero di molecole, ma alle operazioni di una singola macromolecola. Una MM non è una reazione chimica macroscopica [McClare, 1971]. Ciò non toglie che si possa sempre schematizzare una soluzione contenente molte molecole di EcoRI e DNA dicendo che il rapporto tra le molecole con legame specifico e le molecole con legame non specifico è costante all'equilibrio. Tale costante riflette l'energia della reazione (ΔG^0), ma non ci dice nulla sul meccanismo di legame, che è indipendente dalla concentrazione: *una molecola di EcoRI taglia sempre allo stesso modo una molecola di DNA, a prescindere da cosa fanno le altre molecole in soluzione.* Per questo motivo la teoria delle MM non considera le concentrazioni. Come disse McClare (1971), "each molecular machine acts locally as an individual". Si osservi inoltre che EcoRI non è una MM, bensì l'unione di EcoRI e DNA è una MM, infatti il lavoro utile di EcoRI non può essere svolto senza DNA.

Definizione 2.1.3. *Una MM esegue una funzione specifica vantaggiosa per un sistema vivente.*

Oss. 2.1.4. Se la MM non esistesse, la specie ne avrebbe uno svantaggio evolutivo. Shannon sostenne che la teoria dell'informazione non è adeguata a trattare il significato o il valore di una comunicazione [Shannon, 1948; Shannon e Weaver, 1949], ma in biologia è la selezione naturale a fondare i concetti di "funzione" e "utilità".

Definizione 2.1.5. *Una MM è solitamente attivata da una sorgente di energia.*

Oss. 2.1.6. Ciò non significa solamente fotoni e ATP, ma anche energia termica. Ad esempio, è noto che esiste un fenomeno di fotoisomerizzazione termica della rodopsina. E' stato calcolato che l'isomerizzazione spontanea di una molecola di rodopsina per effetto termico dovrebbe verificarsi una volta ogni 3000 anni circa, cioè 10^{23} volte più lentamente rispetto alla fotoisomerizzazione [Baylor, 1987]. L'energia richiesta per la fotoisomerizzazione del retinale è $45kcal/mol$. L'energia associata a una mole di fotoni di lunghezza d'onda $\lambda = 555nm$ (picco di assorbimento della rodopsina in visione fotopica) vale $E = h \frac{c}{\lambda} N_A$ [J], dove $h = 6.6261 \times 10^{-34} Js$ è la costante di Planck, $N_A = 6.0221 \times 10^{23} mole^{-1}$ il numero di Avogadro. Si ottiene $E = 215.7kJ/mole = 51.5kcal/mole$.

Definizione 2.1.7. *Una MM dissipa energia quando esegue un'operazione, vale a dire una sua azione specifica.*

Definizione 2.1.8. Una MM "acquista", o "guadagna" (*gains*) informazione, scegliendo uno tra due o più stati finali nel quale rilassare.

Oss. 2.1.9. EcoRI sceglie tra 4096 esanucleotidi il pattern che corrisponde al suo sito di legame, guadagnando $\log_2(4096) = 12\text{bit}$ di informazione.

Definizione 2.1.10. Le MM operano in condizioni isoterme.

Oss. 2.1.11. Le MM sono costrette a funzionare a $T = \text{cost.}$ in quanto non hanno alcun mezzo di isolamento dal bagno termico che le circonda.

2.2 Esempi di MM

Il DNA può funzionare come una semplice macchina molecolare: la denaturazione di un doppio filamento (*double stranded DNA*, *dsDNA*) produce due filamenti singoli complementari (*single stranded DNA*, *ssDNA*) con energia maggiore; per *ibridizzazione* i due filamenti singoli riformano esattamente la doppia elica originale [Watson e Crick, 1953; Britten e Kohne, 1968]. L'importanza evolutiva della precisione di questo meccanismo è addirittura ovvia.

L'enzima di restrizione *EcoRI* è una proteina che taglia un dsDNA, tra la *G* e la *A* nella sottosequenza *5'GAATTC3'* [Smith, 1979; McClarin e altri, 1986; Rosenberg e altri, 1987b]. *EcoRI* esegue tre operazioni [Rosenberg e altri, 1987a]:

1. Si lega in modo non specifico alla doppia elica (cioè su un punto qualunque).
Nota: da questo istante esiste la MM *EcoRI*+DNA.
2. Scivola casualmente sul filamento fino a che, raggiunta la sottosequenza *5'GAATTC3'*, vi si lega in modo specifico (in questa fase la sua energia libera diminuisce)
3. Esegue il taglio, in presenza di Mg^{2+} . In assenza di Mg^{2+} il legame specifico si instaura ancora, ma non si verifica il taglio.

La prima fase rappresenta l'attivazione della macchina molecolare, in uno stato con contenuto energetico maggiore. Al termine della seconda fase *EcoRI* avrà dissipato energia. Il DNA cellulare è protetto dall'azione di enzimi come *EcoRI*, grazie all'enzima *metilasi*: l'enzima attacca un gruppo metile alla seconda *A* della sequenza *GAATTC*, in modo tale da impedire il taglio. L'enzima metila solo la sequenza in questione e non, ad esempio, *CAATTC*, che pure differisce da *GAATTC* per una sola base [Dugaiczky e altri, 1974]. Quindi, anche se molti esameri simili al sito di legame dell'*EcoRI* restano esposti, *EcoRI* taglia solamente la sequenza giusta [Polisky e altri, 1975; Woodhead e altri, 1981; Pingoud, 1985]. Per molto tempo non si è capito come potesse una singola molecola avere una precisione tale [Rosenberg e altri, 1987a; Needels e altri, 1989; Thielking e altri, 1990]

La *rodopsina* è una proteina contenuta nella retina, capace di rivelare un singolo fotone di luce [Lewis e Priore, 1988; Wessling-Resnick e altri, 1987]. Dopo aver catturato il fotone la rodopsina prima si eccita e poi dissipa energia trasformandosi con alta probabilità in *batorodopsina* (*bathorhodopsin*). La sequenza di reazioni chimiche che segue porta a un guadagno del segnale pari a ≈ 400000 . Si noti che lo stato eccitato della rodopsina è uno stato chimico transitorio, a partire dal quale la molecola ha una certa probabilità di tornare ad essere la rodopsina di partenza. In un certo senso la rodopsina attivata deve compiere una "scelta", in un tempo dell'ordine dei *ps*.

Poco sappiamo sull'esatto funzionamento dei muscoli [Huxley, 1969; McClare, 1971; Highsmith e Jaretzky, 1983; Trayer e Trayer, 1988]. Sappiamo che l'interazione delle proteine *miosina* e *actina* consuma la molecola energetica ATP. L'idrolisi di un ATP porta il complesso *actomiosina* in uno stato eccitato di alta energia, in modo tale che la dissipazione che segue genera una forza.

Altri esempi sono il legame specifico di zuccheri e recettori sulla membrana cellulare [Vyas e altri, 1988], o il legame dell'ossigeno da parte dell'emoglobina [Perutz, 1970].

In tutti gli esempi descritti una certa macromolecola è portata da uno stato iniziale, o stato *before*, in uno stato finale a più bassa energia, o stato *after*. Durante questo passaggio la molecola esegue una operazione specifica. Questa operazione serve a una funzione vantaggiosa, in termini evolutivi, per l'organismo che sintetizza quella macromolecola. Questa operazione può essere caratterizzata in termini del numero di stati distinti tra cui la macchina deve "scegliere" lo stato *after*. Se la macchina può scegliere tra Q stati, diremo che "guadagna" $\log_2 Q$ bit/operazione [Pierce, 1980].

3 Modello di MM

3.1 Assunzioni fondamentali

La configurazione di una MM di n atomi è definita da $3n$ numeri che ne specificano le posizioni rispetto a un sistema di coordinate arbitrario. Se scegliamo come origine il centro di massa, possiamo sottrarre i 6 numeri che definiscono la posizione e l'orientazione del sistema di coordinate sicchè il numero effettivo di coordinate si riduce a $3n - 6$. Altri $3n - 6$ numeri che fissano le velocità degli atomi. $3n - 6$ è il numero di gradi di libertà della MM.

Assunzione 1. Assumiamo che la MM, per funzionare, utilizzi solo alcuni dei suoi gradi di libertà, diciamo

$$d_{space} \ll 3n - 6 \quad (3.1.1)$$

in quanto molti degli atomi costituenti hanno funzione solamente strutturale. Il funzionamento della macchina infatti si basa solitamente sul movimento di precisi aggregati di atomi, che noi chiameremo le sue "parti", o "componenti" (ad esempio, non è necessario seguire tutti gli atomi di una serratura per capire come funziona). La dinamica delle proteine può essere ipotizzata conoscendo il tipo e la posizione dei nuclei degli atomi costituenti. Normalmente le correzioni quantistiche sono trascurabili a 300K [Levy e altri, 1982; Ichiye e Karplus, 1987], ma se occorre considerare anche queste, allora d_{space} sarà il numero dei parametri indipendenti che descrivono il sistema.

Assunzione 2. Assumiamo che le parti di una MM si muovano indipendentemente. Certamente è possibile che il movimento di una parte influenzi il movimento di un'altra parte. Queste correlazioni riducono la dimensionalità dello spazio delle configurazioni, confinando gli stati validi della MM su superfici definite, ma C_y resta comunque un limite superiore per R . Osserviamo che le parti di una MM possono diventare indipendenti attraverso la selezione naturale. Ad esempio, dove è importante che due aminoacidi successivi nella struttura primaria di una proteina possano muoversi indipendentemente (per assolvere alla funzione della proteina), delle mutazioni da inserzione nel gene che codifica quella proteina le conferirebbero un vantaggio selettivo.

Assunzione 3. Assumiamo che l'energetica delle parti di una MM sia descritta dalla *distribuzione di Boltzmann* [Waldram, 1985; Colthup e altri, 1975; Brillouin, 1962]. Ciò equivale ad associare al movimento di ciascuna componente un rumore Gaussiano bianco limitato in banda [Nyquist, 1928; Rice, 1944; Rice, 1945; Shannon, 1949; Petsko e Ringe, 1984]. La stessa cosa è dire che le parti della MM sono in moto Browniano [MacDonald, 1962], dove la velocità è come la somma dell'effetto di tanti piccoli urti. Nelle proteine le fluttuazioni atomiche sono ben caratterizzate da distribuzioni di questo tipo [Ichiye e Karplus, 1987]. Shannon (1949) ha considerato il caso della capacità di canale in presenza di un rumore qualsiasi. Egli notò che tra i molti tipi di rumore esistenti il rumore bianco Gaussiano è in un certo senso il peggiore ai fini della capacità di canale. Nel caso di rumore non Gaussiano gli stati più probabili della MM sono raccolti in forme non sferiche, il che equivale a cambiare la funzione di energia

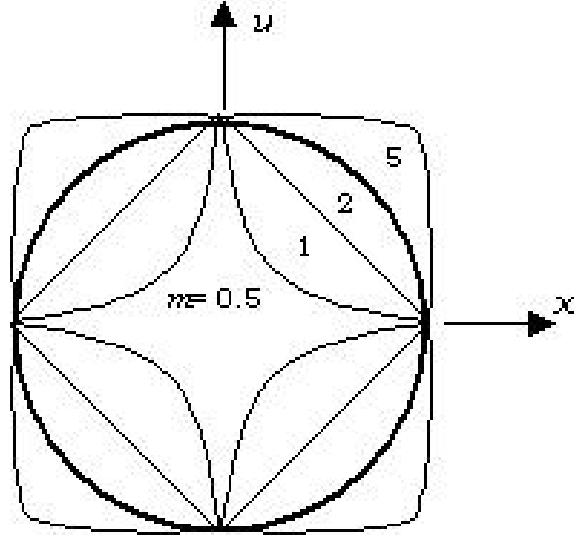


Figure 1: curve $E = cost.$ per diversi valori di m (vedi testo)

delle parti della MM. Ad esempio, se $E_x \propto |x|^m$, $E_y \propto |y|^m$ e $D = 2$, sarebbe $E \propto |x|^m + |y|^m \equiv |r|^m$. Per $m = 2$, $|r|^m = cost.$ descrive un cerchio e il rumore è Gaussiano. Per altri valori di m si ottengono le curve mostrate in Figura 1. Per $m \rightarrow \infty$ la curva tende a un quadrato. Il problema è capire come queste figure, a parità E , possono essere sistemate sul piano nel modo più efficiente, senza sovrapposizioni.

Assunzione 4. Assumiamo che lo stato iniziale sia in equilibrio. La forma della distribuzione degli stati della MM è sferica nello stato finale in quanto ha raggiunto l'equilibrio con il suo ambiente e le parti della MM hanno una distribuzione di Boltzmann. In certi casi anche lo stato iniziale può essere considerato in equilibrio con il suo ambiente, ciò si verifica quando la MM è un sistema fisico "frustrato" [Shakhnovich e Gutin, 1989], come potrebbe essere una molecola di EcoRI su una scala di tempo abbastanza piccola. Questa circostanza fa sì che l'energia sia equidistribuita nei vari modi vibrazionali e rotazionali della MM: l'entropia dello stato *before* è massimizzata e la forma dello stato nello spazio \mathcal{Y} è sferica. Ciò significa che la MM lavora vicino alla sua capacità limite C_y . Ci sono in effetti evidenze che confermerebbero un "complete thermal relaxation" della rodopsina nello stato *before* [Hurley e altri, 1977; Birge e Hubbard, 1980].

3.2 Oscillatore armonico semplice nel vuoto

La posizione e la velocità di un oscillatore armonico libero unidimensionale nel vuoto sono ($d_{space} = 1$; $D = 2$)

$$s(t) = a \cos(\omega t + \varphi) \quad (3.2.1)$$

$$v(t) = \frac{ds}{dt} = -a\omega \sin(\omega t + \varphi) \equiv r \sin(\omega t + \varphi) \quad (3.2.2)$$

$v(t)$ ha due componenti di Fourier [Walker, 1988], x e y

$$v(t) = x \sin(\omega t) + y \cos(\omega t) \quad (3.2.3)$$

$$x \equiv r \cos(\varphi) \quad (3.2.4)$$

$$y \equiv r \sin(\varphi) \quad (3.2.5)$$

Oss. 3.2.1. x e y hanno le dimensioni di una velocità. La coppia (x, y) definisce lo stato del sistema.

L'energia vale [Feynman e altri, 1963]

$$E = \frac{1}{2}mv_m^2 = \frac{1}{2}mr^2 = \frac{1}{2}mx^2 + \frac{1}{2}my^2 \quad (3.2.6)$$

quindi, nel vuoto, dove $E = \text{cost.}$, il luogo di tutti e soli gli stati possibili (x, y) è un cerchio di raggio $r = \sqrt{\frac{2E}{m}}$ nello spazio delle velocità bidimensionale ($D = 2$). Invece di x e y si sarebbero potute considerare le variabili r e ϕ , ma x e y hanno il vantaggio di avere entrambe le dimensioni di una velocità.

3.3 Oscillatore armonico semplice in un bagno termico

Se l'oscillatore è immerso in un bagno termico, gli urti con gli atomi vicini ne cambiano la fase e l'energia, ovvero ϕ e r , in modo casuale. L'equipartizione dell'energia tra oscillatore e il bagno implica che ciascuna componente indipendente di Fourier della velocità ha una distribuzione di Boltzmann [Waldram, 1985]

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{E_x}{2\sigma^2}}, \quad E_x \equiv \frac{1}{2}mx^2 \quad (3.3.1)$$

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{E_y}{2\sigma^2}}, \quad E_y \equiv \frac{1}{2}my^2 \quad (3.3.2)$$

Quindi le velocità x e y hanno distribuzione Gaussiana con deviazione standard proporzionale al parametro σ . Anche il modello classico di Maxwell della distribuzione di velocità delle molecole di un gas ideale usa una distribuzione Gaussiana [wannier, 1966; Castellan, 1971; Waldram, 1985]. La probabilità che l'oscillatore abbia velocità in una piccola regione $dxdy$ è

$$f(x, y)dxdy = f(x)f(y)dxdy = \frac{1}{2\pi\sigma^2} e^{-m\frac{(x^2+y^2)}{4\sigma^2}} dxdy \quad (3.3.3)$$

oppure, in coordinate polari,

$$f(x, y) dx dy = \frac{1}{2\pi\sigma^2} e^{-\frac{mr^2}{4\sigma^2}} dr d\varphi \quad (3.3.4)$$

La probabilità che l'oscillatore abbia velocità in un anello di larghezza dr a distanza r dall'origine vale

$$\int_0^{2\pi} \frac{r}{2\pi\sigma^2} e^{-\frac{mr^2}{4\sigma^2}} dr d\varphi = \frac{r}{\sigma^2} e^{-\frac{mr^2}{4\sigma^2}} dr \quad (3.3.5)$$

$f_R(r) \equiv \frac{r}{\sigma^2} e^{-\frac{mr^2}{4\sigma^2}}$ è la *distribuzione di Rayleigh*. La distribuzione ha simmetria radiale nello spazio (x, y) e tende a zero per $r = 0$.

3.4 MM semplice in un bagno termico

Supponiamo che le energie delle d_{space} parti indipendenti della MM abbiano una distribuzione di Boltzmann, sicché ciascuna di esse si comporta in prima approssimazione come un oscillatore armonico semplice in un bagno termico. $D \equiv 2d_{space}$ numeri servono a descrivere lo stato della macchina (in quanto ciascuna parte ha due componenti di Fourier, x e y , oppure fase φ e ampiezza r). L'energia della parte j -esima è determinata dalla sua velocità e dalla sua massa

$$E_j = \frac{1}{2} m_j v_j^2 \quad (3.4.1)$$

Per confrontare parti di massa diversa ci conviene sostituire alle variabili v_j le variabili $y_j \equiv \sqrt{E_j}$. L'assunzione che le energie abbiano distribuzione di Boltzmann implica che

$$f(y_j) = e^{-\frac{\beta y_j^2}{z}} \quad (3.4.2)$$

con $\beta \equiv \frac{1}{2\sigma^2} = \frac{1}{k_B T}$

essendo definita la *funzione di partizione* $z \equiv \int_{-\infty}^{\infty} e^{-\beta y_j^2} dy_j$

3.5 Spazio Y: MM a molte dimensioni

Disponendo le D v.a. indipendenti y_j lungo assi mutualmente ortogonali individuiamo un singolo punto in uno spazio D -dimensionale. Parafrasando Shannon, *abbiamo sostituito una entità complessa (la configurazione di velocità di una macromolecola) in un ambiente semplice (lo spazio tridimensionale) con una entità semplice (un punto) in un ambiente complicato (spazio D -dimensionale)* [Shannon, 1949; Callen, 1985]. Sia \mathcal{Y} lo spazio delle configurazioni della MM, definito da tutti i possibili valori di y_j . Determiniamo la distribuzione della configurazione della MM alla temperatura T . Lo stato della MM a un dato istante

è dato dall'insieme (y_1, y_2, \dots, y_D) . La densità di probabilità dello stato nello spazio \mathcal{Y} è il prodotto delle densità indipendenti individuali:

$$f(y_1, \dots, y_D) = e^{-\frac{\beta \sum_{j=1}^D y_j^2}{z^D}} = e^{-\frac{\beta \sum_{j=1}^D E_j}{z^D}} = e^{-\frac{\beta N_y}{z^D}} \quad (3.5.1)$$

dove N_y è l'energia del rumore termico, al momento l'unica forma di energia cinetica presente nella macchina. Ma possiamo anche scrivere

$$f(y_1, \dots, y_D) = e^{-\frac{\beta r_y^2}{z^D}} \quad (3.5.2)$$

dove $r_y \equiv \sqrt{\sum_{j=1}^D y_j^2}$ è il raggio medio della sfera D -dimensionale nella quale può trovarsi la configurazione istantanea della MM. Quindi

$$r_y = \sqrt{N_y} \quad (3.5.3)$$

Osserviamo che aver ipotizzato una distribuzione di Boltzmann per le parti della MM implica che esse sono in equilibrio reciprocamente e la MM non dissipa energia. Il rumore termico N_y è lo stesso in qualsiasi configurazione possibile. Quindi r_y è pressoché costante e definisce una sfera nello spazio \mathcal{Y} .

Oss. 3.5.1. La densità di questa sfera tende a concentrarsi vicino alla superficie all'aumentare di D .

Brillouin (1962) dette una prova semplice di questo curioso effetto. Il volume di una sfera D -dimensionale di raggio r è

$$V = \frac{\pi^{\frac{D}{2}} r^D}{\Gamma(\frac{D}{2} + 1)} \quad (3.5.4)$$

dove Γ è la *funzione gamma* [Sommerville, 1929; Shannon, 1949; Kendall, 1961].

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du \quad (3.5.5)$$

Sicché

$$\frac{dV}{V} = D \frac{dr}{r} \quad (3.5.6)$$

Detta $f_D(r)$ la densità di probabilità della sfera, la probabilità che lo stato della MM si trovi in un guscio compreso tra r e $r + dr$ è $f_D(r)dr$, ma anche $f(y_1, \dots, y_D)dV$, se dV è il volume del guscio compreso tra r e $r + dr$. Abbiamo

$$\begin{aligned} f_D(r) &= f(y_1, \dots, y_D) \frac{dV}{dr} = f(y_1, \dots, y_D) D \frac{V}{r} = D \pi^{\frac{D}{2}} r^{D-1} \frac{f(y_1, \dots, y_D)}{\Gamma(\frac{D}{2} + 1)} \\ &= D \pi^{\frac{D}{2}} r^{D-1} \frac{e^{-\beta r_y^2}}{z^D \Gamma(\frac{D}{2} + 1)} \end{aligned} \quad (3.5.7)$$

cioè

$$f_D(r) = \frac{r^{D-1} e^{-\frac{r^2}{2\sigma^2}}}{\Gamma(\frac{D}{2}) \sigma^D 2^{\frac{D}{2}-1}} \quad (3.5.8)$$

La (3.5.8) ha un massimo in $r = \sigma\sqrt{D-1}$.

Se D è abbastanza grande $f_D(r)$ può essere approssimata da una distribuzione Gaussiana.

Oss. 3.5.2. $f_D(r)$ è la funzione di densità di probabilità di una distribuzione χ^2 per la v.a. $x = \frac{r^2}{\sigma^2}$ e D gradi di libertà [Weast e altri, 1988; International Telephone and Telegraph Corporation, 1956].

Oss. 3.5.3. $f_1(r)$ è la *distribuzione Gaussiana*; $f_2(r)$ è la *distribuzione di Rayleigh*; $f_3(r)$ è la *distribuzione di Maxwell-Boltzmann* [Wannier, 1966; Castellan, 1971; Waldram, 1985].

Anche per una piccola molecola, D può essere enorme. P.es. Warshell (1976) ha costruito un modello del retinale 11-cis avente 200 modi vibrazionali e un minimo di $n = 48$ atomi (20 atomi C e 28 atomi H). Quindi $d_{space} = 3n - 6 = 138$, e $D = 2d_{space} = 276$. All'equilibrio termico, se il raggio della sfera aumentasse dell'1%, il volume aumenterebbe del 276%! *Quindi più del 73% del volume della sfera sarebbe concentrato nell'1% dello spessore più esterno!* E in tutto questo abbiamo trascurato l'opsina alla quale il retinale è attaccato, avente p.m. 39048 dalton [Ovchinnikov, 1982; Nathans e Hogness, 1983]. La rodopsina contiene 5511 atomi, quindi $D = 33054$. Certamente non tutti gli atomi sono implicati nel funzionamento della MM, ma il numero di parti che ne definiscono i movimenti resta comunque molto grande.

3.6 Rumore termico nello spazio \mathcal{Y}

Un oscillatore in equilibrio con un bagno termico ha un'energia media $k_B T$ [Waldram, 1985]. Dato che la MM ha d_{space} parti, ciascuna assimilabile a un oscillatore, l'energia totale vale

$$N_y = k_B T d_{space} = \frac{D}{2} k_B T \quad (3.6.1)$$

e ciascuna delle variabili y_j ha un'energia pari a $\frac{k_B}{2T}$. Usando la (3.5.3)

$$r_y = \sqrt{k_B T d_{space}} \quad (3.6.2)$$

Il rumore termico sposta la configurazione della MM lontano dal centro della sfera di una quantità legata alla temperatura assoluta.

3.7 Posizione delle sfere nello spazio \mathcal{Y}

Il quadrato di una distanza nello spazio \mathcal{Y} è sempre pari all'energia che il sistema deve utilizzare per percorrerla. Se rappresentiamo il punto generico dello spazio \mathcal{Y} col vettore D -dimensionale \mathbf{y} , allora possiamo descrivere gli stati istantanei della MM in equilibrio termico con i punti di una sfera

$$\mathbf{y} = \mathbf{s} + \mathbf{N}_y \quad (3.7.1)$$

di cui $\mathbf{s} = (s_1, \dots, s_D)$ è il centro, che definisce la configurazione di velocità media della MM che si trovasse in quello stato di equilibrio, ed $\mathbf{N}_y = (y_1, \dots, y_D)$ il raggio istantaneo. Se la MM passa spontaneamente dalla configurazione media \mathbf{s}_B alla configurazione media \mathbf{s}_A , cede all'ambiente un'energia $|\mathbf{s}_A - \mathbf{s}_B|^2 = \sum_{j=1, \dots, D} (s_{Aj} - s_{Bj})^2$.

3.8 Stato iniziale, stato finale e operazioni di una MM nello spazio \mathcal{Y}

Abbiamo detto che la MM in equilibrio termico con un mezzo a temperatura costante è rappresentabile con una sfera nello spazio \mathcal{Y} . Cosa accade mentre la MM è in funzione? Noi diciamo che una macchina "acquista informazione" quando passa da uno stato più indeterminato a uno stato più determinato. Ora, abbiamo bisogno di considerare solo due stati energetici della MM [Rosenberg e altri, 1987a]. Prima di eseguire un'operazione, nello stato iniziale (anche detto stato *before*), la macchina possiede una certa quantità di energia, mentre dopo, nello stato finale (anche detto stato *after*), ha una quantità di energia inferiore. Lo stato iniziale/finale equivale allo stato del ricevitore di Shannon appena prima/dopo questo abbia determinato un simbolo ricevuto dal canale. P.es. l'operazione di ibridizzazione di due filamenti complementari di DNA in una doppia elica è definita da due stati limite, lo stato iniziale, quando i filamenti sono separati, e lo stato finale, a ibridizzazione avvenuta. Tralasciamo per il momento come faccia la MM a portarsi nello stato iniziale. Fisicamente sarà l'assorbimento di un fotone per la rodopsina [Birge e Hubbard, 1980], o l'idrolisi di ATP per la actomiosina [Weber e Steitz, 1984a], ma nello spazio \mathcal{Y} avremo un'espansione istantanea della sfera che rappresentava lo stato fondamentale della MM in quello che è chiamato stato iniziale. Consideriamo la EcoRI. Quando la molecola è legata al suo sito specifico, nello stato finale, le sue parti hanno energia determinata dal rumore termico. Ogni possibile configurazione della macchina è rappresentata da un punto nello spazio \mathcal{Y} , e l'insieme di tutti questi punti forma una sfera di raggio

$$r_{after} = \sqrt{N_y} \quad (3.8.1)$$

Nello stato iniziale, EcoRI deve avere avuto un'energia interna maggiore di questa, altrimenti non avrebbe potuto attaccarsi al suo sito specifico nello stato finale. Chiamiamo P_y la differenza. Quindi l'energia nello stato iniziale è

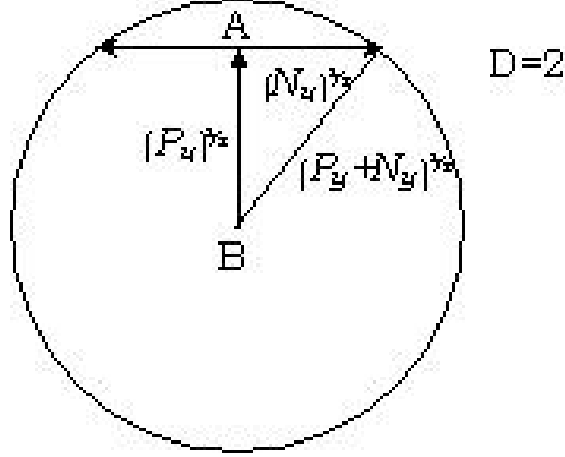


Figure 2: figure title

$P_y + N_y$. Assumiamo che questa energia sia equidistribuita su tutte le parti della MM (assunzione 4). Ciò è ragionevole nel caso della EcoRI, che nel suo stato iniziale si sposta con moto Browniano lungo il DNA. Solo quando EcoRI incontra la sottosequenza *GAATTC* può dissipare l'energia P_y (Assunzione 5). Nello spazio \mathcal{Y} la configurazione della MM nel suo stato iniziale è rappresentata da un vettore di rumore di spostamento dal centro della sfera. L'evento di innesco della MM fornisce a questa un'energia aggiuntiva P_y ; ciò è come sommare al vettore di rumore \mathbf{N}_y un vettore \mathbf{P}_y di ampiezza P_y . Ma in uno spazio con un gran numero di dimensioni la probabilità che \mathbf{P}_y sia ortogonale a \mathbf{N}_y è prossima a uno, sicché \mathbf{N}_y e \mathbf{P}_y formeranno i cateti di un triangolo rettangolo, con ipotenusa di lunghezza $\sqrt{P_y + N_y}$. Dato che tanto \mathbf{N}_y quanto \mathbf{P}_y possono puntare in una direzione qualsiasi, lo stato iniziale è rappresentato da una sfera di raggio

$$r_{before} = \sqrt{P_y + N_y} \quad (3.8.2)$$

avente energia $r_{before}^2 = P_y + N_y$. Durante lo svolgimento dell'operazione, la MM cede un'energia P_y in una data direzione, portandosi nello stato finale rappresentato da una sfera di raggio N_y e centro contenuto nella sfera maggiore dello stato iniziale. Geometricamente \mathbf{P}_y è un vettore con origine nello stato iniziale medio e termine nello stato finale medio (i centri delle rispettive sfere). Non ci preoccupiamo di sapere gli stati di configurazione intermedi.

In Figura la sfera *before* è rappresentata dal cerchio esterno, mentre la sfera *after* dal segmento con centro in A (figura ricavata da Shannon, 1949). Dato che lo spazio \mathcal{Y} ha dimensionalità pari, non è possibile rappresentare visivamente la situazione con $D = 4$. Essendo $D \gg 1$ la gran parte della sfera *after* è appiattita a 90° rispetto alla direzione di P_y . In altre parole la componente di

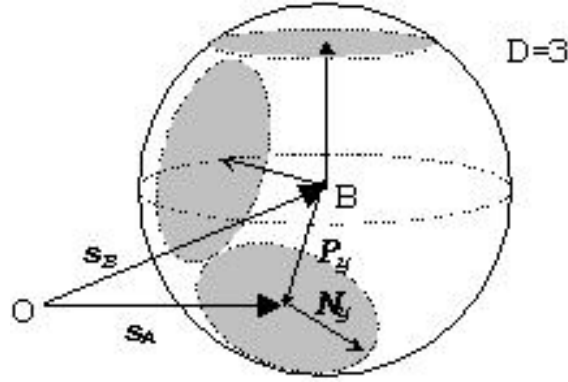


Figure 3: figure title

rumore nella direzione di P_y è sempre più trascurabile all'aumentare del numero di dimensioni D , in quanto il rumore non predilige una direzione in particolare. Il funzionamento della MM corrisponde al movimento del centro della sfera da B ad A, con diminuzione del raggio e cessione dell'energia P_y all'ambiente. Mentre lo stato iniziale è unico, gli stati *after* corrispondenti con una dissipazione P_y sono possibilmente più di uno. Cioè possono esserci diverse sfere di rumore *after* non intersecanti all'interno della sfera *before* [Perutz, 1979; Chothia e Lesk, 1985; Porter e altri, 1983]. La posizione delle sfere *after* in base all'equazione (3.7.1) è detta *schema di codifica* (*coding scheme*), in quanto il raggruppamento delle sfere nello spazio è analogo alla sistemazione delle parole di codifica (*code words*) in un sistema di comunicazione [Sloane, 1984]. Ogni MM ha un suo schema di codifica, e il codice genetico è solo un esempio. La Figura non ha senso fisico, in quanto D è dispari, ma serve solo a illustrare il concetto.

La dimensionalità D determina la definizione del bordo delle sfere, e quindi controlla gli effetti di soglia se due sfere si sovrappongono [Shannon, 1949]. Quindi la precisione di una MM dipende anche dalle sue dimensioni. Se la MM è abbastanza grande ($n \gg 1$) allora il rumore è predicibile in quanto le sfere hanno bordi molto precisi e definiti. Se la MM contiene un numero elevato di parti indipendenti, le sfere possono essere posizionate nello spazio \mathcal{V} in modo che appena evitino il contatto. Perciò la macchina può avere distinti stati *after*. Ma dato che le sfere sono definite da una funzione densità di probabilità del tipo (3.5.8) esse si sovrapporranno sempre in qualche misura, cioè esisterà una probabilità che la MM passi da uno stato finale a un altro stato finale. La probabilità di queste transizioni (la probabilità di transizione errata dallo stato iniziale a uno stato finale errato) è il tasso di errore (*error rate*) della MM. Ovviamente aumentare il numero di atomi per avere D più elevati non garantisce di per sé un posizionamento "intelligente" delle sfere. *L'evoluzione è stata probabilmente l'unica forza che abbia sapientemente guidato il posizionamento delle sfere di tutte le MM.* Sicuramente molti degli aminoacidi che formano la

struttura primaria di una grande proteina possono contribuire in modo fine al posizionamento delle sfere [Needels e altri, 1989].

3.9 Capacità della MM

Quanti possono essere gli stati *after*? Questo numero, \mathcal{M}_y , deve essere minore o uguale al numero massimo di sfere di raggio $\sqrt{N_y}$ contenute in una sfera di raggio $\sqrt{P_y + N_y}$ [Sloane, 1984; Conway e Sloane, 1988]

$$\mathcal{M}_y \leq \frac{V_{before}}{V_{after}} = \left(\frac{P_y + N_y}{N_y} \right)^D \quad (3.9.1)$$

Definizione 3.9.1. La *capacità di macchina* è l'informazione massima che può essere acquistata dalla MM durante l'operazione [Shannon, 1948; Shannon e Weaver, 1949; Shannon, 1949; Pierce, 1980].

$$\mathcal{C}_y \equiv \log_2 \mathcal{M}_y = d_{space} \log_2 \left(1 + \frac{P_y}{N_y} \right) \quad [bit/operazione] \quad (3.9.2)$$

Oss. 3.9.2. L'assunzione che le parti della MM siano indipendenti implica che essa sia schematizzabile come un gas ideale nello spazio \mathcal{Y} , e il funzionamento della MM come il collassamento di questo gas. In base alla termodinamica classica, la diminuzione di entropia è pari al logaritmo del rapporto dei volumi, quindi è uguale all'acquisto di informazione \mathcal{C}_y .

Oss. 3.9.3. La formula per \mathcal{C}_y vale in presenza di rumore Gaussiano. In presenza di altri tipi di rumore essa rappresenta comunque un limite superiore che non può essere superato.

Oss. 3.9.4. La dimostrazione del teorema che esprime la capacità di canale è basata su semplici concetti geometrici.

Esempio. La *trascrizione*, che avviene nel nucleo, è la sintesi di una copia del filamento di DNA di un gene, catalizzata dall'enzima *RNA-polimerasi*. Questa copia è detta *trascritto primario*, o *mRNA*. Supponiamo che a 300 K una RNA-polimerasi inserisca le basi (*A, G, C, U*) con una velocità di $W_S = 200$ operazioni/sec [Golomb e Chamberlin, 1974]. Ad ogni operazione essa inserisce una base e acquista $R = \log_2 4 = 2$ bits. Possiamo considerare $W_S R = 400$ bit/sec la sua capacità. Supponiamo ora che la temperatura aumenti, e con essa il rumore termico. Le sfere *after* tendono a espandersi e quindi a sovrapporsi. Supponiamo che a causa di questa sovrapposizione gli stati *after* *A* e *G* divengano indistinguibili, cioè tendano a formare un unico sotto-stato, e così pure *C* e *U*, mentre W_S non cambi significativamente a seguito dell'aumento di *T*. In queste condizioni la RNA-polimerasi ottiene solo 1 bit/operazione, e la sua capacità si riduce a 200 bit/sec.

3.10 Codifica di una MM

E' vantaggioso per un sistema biologico trovare il modo di "impacchettare" le sfere *after* in poco spazio, infatti meno energia dovrà essere dissipata per ogni operazione, in quanto, a parità di informazione acquisita per operazione, la sfera *before* sarà più piccola. *Il problema della sistemazione delle sfere equivale a trovare un codice di comunicazione ottimo.* Ciascuna MM ha il proprio codice. Nel modello di Shannon più è complessa la codifica del segnale (più numeri uso per descrivere il segnale), maggiore è la dimensionalità dello spazio \mathcal{Y} , più le sfere hanno frontiere definite e dense, minore è la probabilità di intersezione e quindi di errore. Chiaramente codifiche più laboriose richiedono un tempo maggiore, ma ci danno la possibilità di approssimare la capacità del canale con un tasso di errore accettabile! Una MM *semplice*, come quelle che abbiamo considerato finora, può ottenere lo stesso scopo solo attraverso il movimento delle sue parti nello spazio, in quanto la sua profondità di memoria temporale è 1. In altre parole le MM tendono a usare una codifica spaziale piuttosto che una codifica temporale. Di fatto, tutte o quasi tutte le MM descritte dai biologi molecolari prevedono meccanismi allosterici [Monod e altri, 1965] e interazioni del tipo "chiave-serratura" [Rastetter, 1983; Gilbert e Greenberg, 1984]. Possiamo quindi enunciare il teorema di Shannon (applicato alle MM):

aumentando il numero di parti in movimento indipendente che possono interagire in cooperazione per prendere decisioni, una MM può ridurre la frequenza di errore a un livello tanto piccolo quanto è richiesto ai fini della sopravvivenza dell'organismo, anche quando la MM operi vicino alla sua capacità e dissipi piccole quantità di energia.

Nella teoria della comunicazione un ricevitore è un dispositivo il cui stato è determinato da un segnale esterno variabile nel tempo, opportunamente decodificato. Ma una MM semplice, come EcoRI, non viene portata nel suo stato finale (il sito di restrizione) da un comando esterno. Inoltre la decodifica di un segnale variabile nel tempo (come la codifica) richiede un *buffer* di memoria dove il segnale stesso possa essere registrato. Ma una MM semplice non possiede memoria. Invece il DNA nel canale di EcoRI agisce come una chiave nella serratura, dove il processo di riconoscimento si attua in parallelo sulla superficie di contatto tra DNA e EcoRI [Rastetter, 1983; Gilbert e Greenberg, 1984; McClarin e altri; 1986]. Tuttavia anche un messaggio codificato nel tempo può essere ricevuto, immagazzinato ed elaborato da una combinazione di più MM semplici. Questa sorta di *ricevitori molecolari* (o *MR*) potrebbero decodificare un messaggio, e funzionare da interfaccia tra l'uomo e un computer molecolare. Non è noto se ogni organismo vivente possieda dispositivi simili, ma possibili candidati sono i processi di traduzione, il movimento cellulare, la mitosi, lo sviluppo embrionale, i ritmi circadiani, ecc., cioè tutti i processi biologici guidati da segnali che evolvono nel tempo.

In base all'analisi di Fourier, un segnale variabile nel tempo può essere registrato sotto forma di una serie di campioni. Se B è la frequenza superiore

dello spettro del segnale e τ la sua durata, allora il segnale stesso può essere ricostruito fedelmente avendo almeno $d_{time} \equiv 2\tau B$ campioni [Shannon, 1949; Conway e Sloane, 1988; Walker, 1988]. Se i diversi stati del MR sono determinati da un segnale esterno, la sua codifica può essere descritta in uno spazio \mathcal{Z} a $D = d_{space}d_{time}$ dimensioni. L'energia totale media del MR è

$$\langle E_z \rangle = \frac{k_B T}{2D} = \tau B k_B T d_{space} \quad [J] \quad (3.10.1)$$

Il rumore termico totale è

$$N_z \equiv \frac{\langle E_z \rangle}{\tau} = B k_B T d_{space} \quad [J/s] \quad (3.10.2)$$

La densità di probabilità della sfera è ancora data dalla (3.5.8). Il volume della sfera è proporzionale al raggio elevato alla dimensionalità dello spazio che contiene la sfera stessa, sicché il massimo numero di stati *after* contenibili dallo stato iniziale è

$$\mathcal{M}_y \leq V_{before}/V_{after} = D \frac{P_z + N_z}{N_z} \quad (3.10.3)$$

Definizione 3.10.1. La **capacità del ricevitore molecolare** è quindi

$$\mathcal{C}_z \equiv \log_2\left(\frac{\mathcal{M}_y}{\tau}\right) = d_{space} B \log_2\left(1 + \frac{P_z}{N_z}\right) \quad [bit/sec] \quad (3.10.4)$$

Questa espressione è più generale di quella della capacità di macchina \mathcal{C}_y , in quanto contempla una codifica/decodifica sia nel tempo che nello spazio. Se $d_{space} = 1$ c'è un solo grado di libertà spaziale, e si ottiene la formula di Shannon, mentre l'espressione di N_z si riduce a quella di Nyquist [Nyquist, 1928; Johnson, 1928; Pierce, 1980]. \mathcal{C}_z può essere vista come la capacità di d_{space} canali di comunicazione in parallelo.

4 Dissipazione di una MM

4.1 Introduzione

Si dimostra che la quantità minima di energia che la MM deve dissipare per acquisire un bit di informazione (cioè prendere una decisione binaria) è $\epsilon_{min} = k_B \ln(2)$. Consideriamo l'ibridizzazione del DNA. Essa è definita da due stati limite, lo stato iniziale quando i due filamenti sono separati, e lo stato finale, ad ibridizzazione avvenuta. Per la prima legge della termodinamica

$$\Delta U = q + w \quad (4.1.1)$$

cioè l'energia interna varia di una quantità pari alla somma algebrica del calore assorbito dalla MM e del lavoro fatto dalla MM sull'ambiente. Ma una molecola di DNA non compie lavoro su un oggetto esterno mentre si ibridizza, sicché $w = 0$. Dato che $\Delta U < 0$, dev'essere $q < 0$, cioè calore viene dissipato all'esterno. Come possiamo allora quantificare l'azione fatta dalla MM se non ha prodotto lavoro? L'aspetto importante dal punto di vista biologico è il numero di scelte che la MM compie. Per formare ciascuna bp (*base pair*) di DNA, solo 4 possibilità su 16 sono accettabili (se A,C,G,T allora T,G,C,A, rispettivamente). Quindi la MM deve acquisire $\log_2 4 = 2$ bits di informazione per ogni bp.

4.2 Incertezza, entropia, informazione

Ciascuna configurazione possibile della MM corrisponde a un particolare microstato. Se Ω è il numero totale di microstati, ciascuno con probabilità $P_i \geq 0$,

$$\sum_{i=1}^{\Omega} P_i = 1 \quad (4.2.1)$$

L'insieme di tutti i microstati forma una sfera in uno spazio a molte dimensioni [Schneider, 1991], e Ω è proporzionale al volume di questa sfera [Callen, 1985]. La misura dell'informazione secondo Shannon è legata al concetto di **incertezza** (*uncertainty*) dei microstati. Secondo la formula di Shannon tale incertezza vale

$$H \equiv - \sum_{i=1}^{\Omega} P_i \log_2 P_i \quad [bit/microstato] \quad (4.2.2)$$

L'**entropia statistica** di Boltzmann-Gibbs di un sistema fisico, quale è la MM, vale

$$S \equiv -k_B \sum_{i=1}^{\Omega} P_i \ln P_i \quad [JK^{-1}microstato^{-1}] \quad (4.2.3)$$

dove $k_B = 1.38 \times 10^{-23} JK^{-1}$. Dato che $\log_2 x = \frac{\ln x}{\ln 2}$

$$S = k_B \ln(2)H \quad (4.2.4)$$

Quindi l'unica differenza tra entropia e incertezza è l'unità di misura [von Neumann, 1963; Brillouin, 1962; Rothstein, 1951]. L'entropia di una MM può diminuire a spese di un aumento di entropia nell'ambiente circostante. La riduzione di entropia durante l'operazione di una MM

$$\Delta S = S_{after} - S_{before} \quad [JK^{-1}operazione^{-1}] \quad (4.2.5)$$

corrisponde a una diminuzione dell'incertezza

$$\Delta H = H_{after} - H_{before} \quad [bit/operazione] \quad (4.2.6)$$

e

$$\Delta S = k_B \ln(2)\Delta H \quad (4.2.7)$$

La diminuzione di incertezza corrisponde a un guadagno di informazione R , definito [Shannon, 1948; Brillouin, 1951b; Tribus e McIrvine, 1971; Rothstein, 1951]

$$R \equiv -\Delta H \quad [bit/operazione] \quad (4.2.8)$$

Se si conosce il numero di operazioni svolte in un secondo, R può immediatamente essere convertito in bit/sec. Definita in questo modo, l'informazione è acquistata solo quando la MM passa da uno stato più indeterminato a uno più determinato [Schneider, 1991].

Oss. 4.2.1. Non sempre $H_{after} = 0$. Per esempio, si consideri il caso di un riconoscitore di una sottosequenza di DNA che può accettare indifferentemente una purina in una certa posizione. In tal caso $H_{after} = 1$ bit (infatti la base può essere A oppure G). Per questo motivo non è sempre corretto eguagliare R con H_{before} . Questo punto ha spesso ingenerato confusione [Popper, 1967a; Popper 1967b; Wilson, 1968; Ryan, 1972; Ryan, 1975].

R è massimizzato quando H_{after} è minimo e H_{before} è massimo. Quest'ultima condizione si verifica se i microstati sono equiprobabili, cioè se $H \equiv \log_2 \Omega$, cioè $P_i \equiv \Omega^{-1}$. Infine

$$\Delta S = -k_B \ln(2)R \quad (4.2.9)$$

Questa equazione collega la diminuzione di entropia e l'acquisto di informazione che una MM ottiene nell'eseguire un'operazione [Rothstein, 1951]. In altre parole perché una MM possa eseguire delle scelte la sua entropia deve diminuire. Notiamo che in questo caso ΔS rappresenta solo quella parte della

variazione totale di entropia dovuta alla selezione dei microstati che la MM compie durante la sua evoluzione. $R > 0$ e $\Delta S < 0$. L'entropia totale dell'universo è comunque non decrescente, cioè

$$\Delta S_{universe} = \Delta S + \Delta S_{ambiente} \quad (4.2.10)$$

P.es. l'uguaglianza vale quando una soluzione di EcoRI e DNA è in equilibrio in assenza di Mg^{2+} . Durante lo svolgimento di un'operazione l'entropia dell'ambiente che localmente circonda la MM, cioè il fluido in cui la MM si trova, deve aumentare. Per la seconda legge della Termodinamica

$$dS \geq \frac{dq}{T} \quad (4.2.11)$$

dove l'uguaglianza vale nel caso di trasformazioni reversibili. Quindi, considerato un piccolo volume che contenga esattamente la MM, se il calore dq è assorbito da questo volumetto, allora l'entropia all'interno deve aumentare almeno di $\frac{dq}{T}$ [Tolman, 1938; Fermi, 1936; Jaynes, 1988]. Le MM sono completamente esposte al solvente, sicché la loro temperatura è la stessa nello stato iniziale e nello stato finale, cioè le MM funzionano sempre in condizioni isoterme ($T = cost.$). Integrando abbiamo

$$\Delta S \geq \frac{q}{T} \quad (4.2.12)$$

da cui

$$-\frac{q}{R} \geq k_B \ln(2) \quad [J/bit] \quad (4.2.13)$$

Esiste pertanto una energia minima

$$\epsilon_{min} = k_B \ln(2) \quad [J/bit] \quad (4.2.14)$$

che la MM deve dissipare affinché possa acquistare $R = 1$ bit di informazione. In altre parole

il limite superiore all'informazione che può essere guadagnata con precisione illimitata da una MM per 1 J di energia dissipata è $\frac{1}{k_B \ln(2)}$ bits.

Questa è una conseguenza diretta del secondo principio della termodinamica in condizioni isoterme. E' molto probabile che la derivazione data sia la stessa seguita da von Neumann durante le sue lezioni all'Università del Illinois nel 1949 (von Neumann, 1966). Per ironia della sorte le sue parole esatte sono andate perdute a causa del rumore sul nastro di registrazione.

Lo stesso risultato può essere derivato a partire dalla formula della capacità di macchina C_y . Questa derivazione fu ideata per primo da Pierce e Cutler [Pierce e Cutler, 1959; Raisbeck, 1963]. Quello che ci stupisce è che per moltissimo tempo si è ignorato lo stretto collegamento che esiste tra il secondo

principio della termodinamica e la capacità di canale ricavata da Shannon nel 1949 [Jaynes, 1988; Ehrenberg, 1967]. In questo caso è $P_y \equiv -q$. Si ha

$$\epsilon_{min} \equiv \lim_{P_y \rightarrow 0} \frac{P_y}{C_y} = \lim_{P_y \rightarrow 0} \frac{P_y}{d_{space} \log_2(1 + \frac{P_y}{N_y})} \quad (4.2.15)$$

Per il teorema di l'Hopital possiamo derivare numeratore e denominatore rispetto a P_y [Thomas, 1968], ottenendo

$$\epsilon_{min} = \lim_{P_y \rightarrow 0} \ln(2) \frac{N_y}{d_{space}} \quad (4.2.16)$$

Per N_y possiamo usare la formula di Nyquist per il rumore termico [H.Nyquist. Thermal agitation of electric charge in conductors. Physical Review, 32:110-113, 1928]. Per ciascun grado di libertà il rumore termico vale $\frac{1}{2}k_B T$. Ciascuna delle d_{space} parti della MM contribuisce con due gradi di libertà. Quindi, usando la (3.6.1) si riottiene la (4.2.14)

$$\epsilon_{min} = k_B \ln(2) \quad [\text{J/bit}]$$

Questo risultato è del tutto generale, e vale anche per un ricevitore di Shannon [Shannon, 1949] o un *ricevitore molecolare* [Schneider, 1991]. In quest'ultimo caso, sempre con $T = cost.$,

$$\epsilon_{min} \equiv \lim_{P_z \rightarrow 0} \frac{P_z}{C_z} \quad (4.2.17)$$

ma il limite $P_z \equiv \frac{-q}{t} \rightarrow 0$ può essere ottenuto in due modi: o per t finito e $-q \rightarrow 0$ (allo stesso modo in cui ciò avviene per una MM semplice allorché $P_y \equiv -q \rightarrow 0$), oppure per $-q$ finito e $t \rightarrow \infty$, che equivale a una trasformazione isoterma reversibile di durata infinita. Quest'ultimo caso non si verifica nelle MM semplici, che non hanno alcun modo di sfruttare a loro vantaggio la variabile tempo.

Oss. 4.2.2. Osserviamo che la MM potrebbe dissipare più di ϵ_{min} durante un'operazione, ma certamente non di meno.

5 Computazione tramite MM

Tutte le MM eseguono operazioni logiche. P.es. se un *ssDNA* contenente 5'TAC3' ibridizza, allora il filamento associato deve contenere

(A in pos.1) **AND** (T in pos.2) **AND** (G in pos.3)

Analogamente EcoRI taglia il DNA solo nel sito definito da

(G in pos.1) **AND** (A in pos.2) **AND** (A in pos.3) **AND** (T in pos.4) **AND** (T in pos.5) **AND** (C in pos.6)

Possiamo ipotizzare un altro enzima di restrizione con sito

(C **OR** G in pos.1) **AND** (A in pos.2) **AND** (T in pos.3) **AND** (T in pos.4)

[Smith, 1979; Roberts, 1989]. Oppure, la proteina repressore *lac* si lega all'operatore solo se non (**NOT**) è a sua volta legata a un induttore. Come è noto qualsiasi funzione aritmetica e logica può essere ottenuta combinando **AND** e **NOT** [Wait, 1967; Gersting, 1986; Schilling e altri, 1989].

Secondo Bennett (1987) e Landauer (1988) non occorre dissipare energia per eseguire operazioni logiche e aritmetiche. In effetti EcoRI esegue, senza dissipare energia, un'operazione Booleana ogniquale volta si lega a un *dsDNA* in modo non specifico (cioè non sul sito di restrizione). In questo caso potremmo dire che il risultato dell'operazione è sempre **FALSE**. Quando invece EcoRI incontra il sito di restrizione il risultato è **TRUE**, ma stavolta l'energia dell'enzima si abbassa. Pertanto il completamento di un'operazione da parte di una MM in presenza di rumore, comporta sempre una dissipazione di energia. Cioè, come Feynman precisò (1987), anche se la computazione in sé può non essere associata a dissipazione di energia verso l'esterno, l'emissione del risultato (output) lo è sempre. Il quantitativo minimo di energia richiesto per emettere un output binario è $\epsilon_{min} = k_B \ln(2) \text{ J/bit}$. Anche una singola molecola può eseguire una operazione logica Booleana con precisione illimitata, purché non ecceda la sua capacità di macchina. Questo ha convinto molti che i computers molecolari sono una possibilità reale [Feynman, 1961; Drexler, 1981; Carter, 1984; Haddon e Lamola, 1985; Conrad, 1985; Conrad, 1986; Drexler, 1986; Arrhenius e altri, 1986; Hong, 1986; Hopfield e altri, 1988; Eigler e Schweizer, 1990]. Anche se attualmente non sappiamo come ottenere computers fatti di proteine le capacità tecnologiche per farlo già esistono [Maniatis e altri, 1982; Beaucage e Caruthers, 1981; Pabo, 1983; Ulmer, 1983; Rastetter, 1983; Wetzel, 1986; Lesk, 1988].

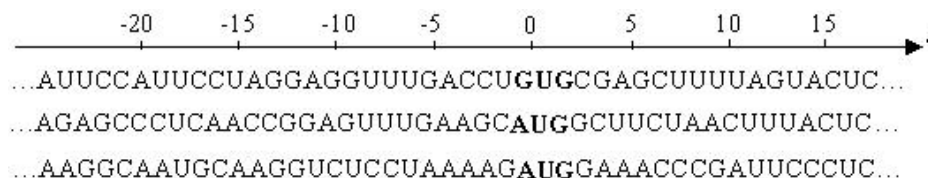


Figure 4: figure title

6 Informazione biologica nei siti di legame

6.1 Informazione biologica nel sito di legame ribosomale di *E. coli*

Un *ribosoma* è un insieme di proteine e RNA che legge mRNA e usa questa informazione per costruire altre proteine. Questo processo, detto **traduzione** (*translation*), inizia da una regione del trascritto primario detta **sito di legame ribosomale** [Gold e altri, 1981; Stormo e altri, 1982b]. Uno dei problemi che un ribosoma deve risolvere è trovare il sito di legame. La sopravvivenza della cellula dipende da quanto efficiente sia questa ricerca. Se il ribosoma iniziasse la traduzione dal punto sbagliato la cellula consumerebbe energia per produrre proteine inservibili. Il batterio *Escherichia coli* ha circa 2600 geni [Schneider e altri, 1986]. Questi geni devono essere localizzati, o selezionati, su un totale di 4.7×10^6 basi che formano il trascritto primario dell'intero genoma [Kohara e altri, 1987]. Quindi il problema è quello di localizzare $\gamma = 2600$ oggetti su $G = 4.7 \times 10^6$ oggetti (la base iniziale di ciascun gene sul totale delle basi possibili), senza commettere errori. Durante la ricerca il ribosoma si sposta con un moto casuale Browniano. Per ciascuna operazione di ricerca il numero di scelte binarie che devono essere fatte è

$$R_{frequency} \equiv \log_2 \left(\frac{G}{\gamma} \right) = 10.8 \quad [bit/operazione] \quad (6.1.1)$$

Rispetto alla scala temporale delle mutazioni genetiche sui siti di inizio traduzione, la dimensione G dell'intero genoma è approssimativamente costante. Analogamente, il numero di siti γ è fissato dalle funzioni fisiologiche che sono sotto il controllo del riconoscitore. Dato che $R_{frequency}$ dipende dalla dimensione del genoma e dal numero di geni funzionali, ne rappresenta una caratteristica specifica e abbastanza costante su lunghi periodi di tempo. Il ribosoma inizia la traduzione da un cluster di G e A, seguito da un intervallo di una decina di basi, quindi da un *AUG* (oppure *GUG*, *UUG*). La figura seguente mostra tre dei siti di legame ribosomali di *E. coli* [Gold e altri, 1981; Stormo e altri, 1982b].

Se disponiamo su righe successive un certo numero di queste sequenze, in modo tale che le partenze dei codoni di inizio si trovino tutte allineate nella stessa colonna (diciamo, ascissa $l = 0$), possiamo calcolare la frequenza della

base b in posizione l

$$f(b, l) : \{A, C, G, U\} \rightarrow \{\dots - 2, -1, 0, 1, 2, \dots\} \quad (6.1.2)$$

Infatti

$$f(b, l) \equiv \frac{n(b, l)}{n(l)} \quad (6.1.3)$$

dove $n(b, l)$ è il numero di basi b che si trovano in posizione l , e $n(l)$ è il numero di sequenze complete della base l (talvolta possono esserci sequenze frammentate)

$$n(l) \equiv \sum_{b \in \{A, G, C, T\}} n(b, l) \quad (6.1.4)$$

Ciascuno dei punti sensibili del ribosoma, sfiora il trascritto primario in una determinata posizione e tenta di identificare la base. A una data posizione l , l'incertezza a priori (*before*), cioè prima del riconoscimento, è $H_{before} = \log_2 4 = 2$ bit/base, in quanto tutte e 4 le basi sono ugualmente possibili, mentre l'incertezza a posteriori (*after*) è [Shannon, 1948; Pierce, 1980; Schneider e altri, 1986; Schneider, 1988; Schneider e Stephens, 1990; Stephens e Schneider, 1992; Shaner e altri, 1993]

$$H_{after}(l) \equiv - \sum_{b \in \{A, G, C, T\}} f(b, l) \log_2 f(b, l) + e(n(l)) \quad [bit/base] \quad (6.1.5)$$

Chiaramente l'accuratezza della $H_{after}(l)$ dipende dalla numerosità delle sequenze utilizzate, $n(l)$. Infatti secondo la teoria canonica dell'informazione l'incertezza dovrebbe dipendere dalle probabilità di occorrenza delle basi, mentre invece noi usiamo delle frequenze. Per questo motivo compare il termine di correzione $e(n)$, spesso trascurabile. Il contenuto di informazione che il ribosoma trova in prossimità del sito di legame è uguale alla diminuzione dell'incertezza passando dallo stato iniziale allo stato finale, quindi

$$\begin{aligned} R_{sequence}(l) &\equiv -\Delta H = H_{before}(l) - H_{after}(l) = \\ &= 2 + \sum_{b \in \{A, G, C, T\}} f(b, l) \log_2 f(b, l) \quad [bit/base] \end{aligned} \quad (6.1.6)$$

Spesso si lavora con il logo della sequenza (sequence logo), che mostra su ciascuna ascissa l le lettere delle basi $b \in \{A, C, G, U\}$ impilate l'una sopra l'altra, aventi altezza $\propto f(b, l)$ e altezza cumulativa pari a $R_{sequence}(l)$.

La figura 6.1 mostra il logo della sequenza ricavato da $n(l) = 149$ siti di legame ribosomale di *E.coli* [Schneider e Stephens, 1990]. Le barre di errore si riferiscono all'altezza cumulativa. Dato che ogni base incontrata è una sorgente indipendente di informazione (ciò è generalmente vero, ma può comunque essere

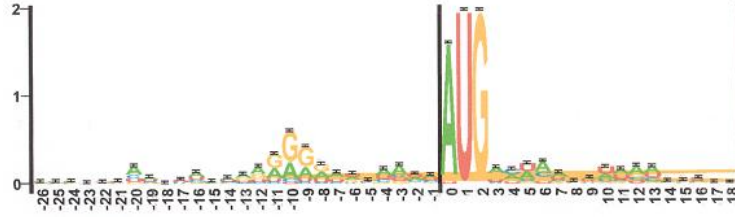


Figure 5: sito di legame ribosomale di *E.coli*

verificato), per la proprietà dei logaritmi l'informazione sull'intero pattern del sito di legame si ottiene sommando $R_{sequence}(l)$ su tutte le posizioni

$$R_{sequence} \equiv \sum_l R_{sequence}(l) \quad [bit/sito] \quad (6.1.7)$$

Si ricava $R_{sequence} = 11.0 \pm 0.4$ bits, valore molto simile a $R_{frequency} = 10.8$ bits. Questo significa che la quantità di informazione presente nel pattern del sito di legame è sufficiente per la individuazione dei siti stessi all'interno del genoma della cellula. Inoltre è evidente che $R_{sequence}$ deve evolvere, per selezione naturale, verso $R_{frequency}$!

6.2 Informazione biologica nel sito accettore di splice dell'uomo

Nelle nostre cellule il DNA viene copiato in filamenti di RNA. Da questo RNA il processo di splicing rimuove alcuni segmenti non codificanti (introni) tramite dei tagli effettuati in punti specifici. I siti di taglio che delimitano un introne sono detti sito donore (verso l'estremità 5') e sito accettore (verso l'estremità 3'). Nell'uomo, per i siti accettori si ricava $R_{sequence} \simeq 9.4$ bits. La lunghezza del genoma umano, $G \approx 4 \times 10^9$ basi, divisa per il numero g di siti accettori è pari alla distanza media tra siti accettori, cioè alla lunghezza media degli introni più quella degli esoni. Questa quantità vale $\simeq 812$ basi, quindi l'informazione necessaria per rintracciare tutti i siti accettori di splice è $R_{frequency} \equiv \log_2(\frac{G}{g}) = \log_2 812 = 9.7$ bit. Ancora una volta $R_{sequence}$ è molto prossimo a $R_{frequency}$! La Figura seguente mostra il logo della sequenza ottenuta da Mike Stephens (Linganore High School) raccogliendo 1744 siti accettori di splice umani [Journal of Molecular Biology 228: 1124-1136, 1992).

6.3 Informazione biologica nei promotori del batteriofago T7

Nel caso dei promotori del virus batteriofago T7 si incontra una deviazione dalla regola $R_{sequence} \approx R_{frequency}$. T7 infetta il batterio *E. coli* e sostituisce la sua RNA-polimerasi a quella della cellula ospite. Raccogliendo sequenze dei siti

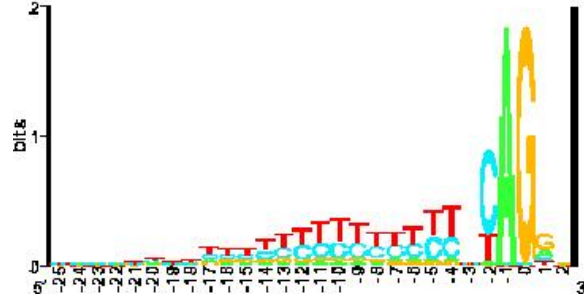


Figure 6: figure title

riconosciuti da queste RNA-polimerasi si è trovato $R_{sequence} = 35.4 \pm 2.0$ bit. La cosa strana è che la quantità di informazione che serve per identificare tutti questi siti è solamente $R_{frequency} = 16.5$ bit. Quindi $R_{sequence} \simeq 2R_{frequency}$ [Schneider, Stormo, Gold e Ehrenfeucht; 1986]. Una spiegazione è che nel sito coesistano due codici per due proteine differenti e non interferenti che si legano in modo specifico ai promotori di T7, probabilmente usando contatti molecolari diversi. Una di queste proteine è sicuramente la T7 RNA-polimerasi. Anche nel caso dei siti *incD* del plasmide F si trova $R_{sequence} = 60.2 \pm 2.6$ bits, $R_{frequency} = 19.6$ bits. Stavolta $R_{sequence} \simeq 3R_{frequency}$. In effetti si è trovato che esistono tre proteine non interferenti che si legano selettivamente a questo sito [N.D.Herman e T.D.Schneider, 1992].

6.4 Oscillazioni nella distribuzione di informazione nelle sequenze di DNA

La combinazione di vari fattori come la relativa rigidità dei filamenti singoli e la natura debole dei legami idrogeno fa sì che le due eliche del DNA non siano diametralmente opposte rispetto all'asse molecolare, ma che esista uno sfasamento. Ragion per cui il B-DNA presenta un solco maggiore e un solco minore (major and minor grooves). Ciò ha delle implicazioni biochimiche, tanto è vero che esistono proteine che si legano di preferenza al solco maggiore. Ora, si osserva che l'informazione $R_{frequency}(l)$ presenta spesso due picchi nel pattern del sito di legame, distanziati di una decina di basi. Si pensa che ciò sia dovuto proprio alla presenza del solco maggiore e del solco minore. I picchi di informazione si troverebbero in corrispondenza dei punti nei quali la proteina affronta il solco maggiore (si ricordi che la doppia elica compie una rotazione completa in 10 basi circa). Il solco maggiore è più ampio del solco minore e permetterebbe alla proteina di riconoscere tutte e quattro le configurazioni *AT*, *TA*, *CG*, *GC*. Invece il solco minore non lascia distinguere *AT* da *TA*, né *CG* da *GC*. Quindi la MM può acquisire fino a 2 bit di informazione per posizione in corrispondenza di un solco maggiore, ma solo 1 bit in corrispondenza di un solco minore. Di

fatto in qualche caso $R_{frequency}(l)$ oscilla proprio tra i valori 2 bit e 1 bit.

7 Appendice I

I biologi molecolari scoprono continuamente esempi di macchine molecolari, e molti scienziati hanno discusso i potenziali vantaggi di poterle costruire (Feynman, 1961; McClare, 1971; Drexler, 1981; Carter, 1984; Haddon e Lamola, 1985; Conrad, 1985; Conrad, 1986; Drexler, 1986; Arrhenius e altri, 1986; Hong, 1986; Hopfield e altri, 1988).

Di seguito introduciamo le principali nozioni riguardo ad alcuni importanti sistemi biologici.

7.1 trasduzione Fotochimica

Le membrane dei dischi del segmento esterno dei coni e dei bastoncelli contengono i fotopigmenti visivi, in grado di catturare la luce. L'orientamento delle molecole del fotopigmento è tale da massimizzare la probabilità di assorbire i fotoni incidenti. I fotopigmenti sono formati da un'aldeide della vitamina A (*cromoforo*) e da una proteina, l'*opsina*, di cui vi sono tre varietà nei coni e una varietà nei bastoncelli.

$$\text{fotopigmento} = \text{cromoforo} + \text{opsina}$$

L'opsina e il cromoforo sono uniti da una base di Schiff. I pigmenti sono la *rodopsina* nei bastoncelli e la *iodopsina* nei coni. Nel segmento esterno di ogni bastoncello umano ci sono $10^8 - 10^9$ molecole di rodopsina disposte nei dischi in strati monomolecolari. Esse rappresentano l'80% delle proteine della membrana e la distanza tra le molecole è di soli 20 nm. Il gruppo cromoforo della rodopsina è il *retinale*, un lipide a lunga catena derivato dal retinolo (vitamina A).

$$\text{rodopsina} = \text{retinale} + \text{opsina}$$

I retinali sono molecole formate da una catena coniugata di carbonio, terminata con un anello (*b-ionile*) da una parte e con un atomo di ossigeno dall'altra. Il retinale può assumere due diverse conformazioni isomeriche: l'isomero *11-cis* (a catena ripiegata) che può legarsi a una lisina dell'opsina attraverso una base di Schiff protonata, e l'isomero *tutto-trans* (*all-trans*, a catena dritta), che non può legarsi all'opsina. Quando una molecola di rodopsina assorbe un fotone il retinale assume la forma tutto-trans, per rotazione della catena terminale connessa all'opsina, e si scinde dall'opsina (*fotoisomerizzazione*). Questo processo dura $\approx 10^{-12}$ sec.

Subito dopo aver assorbito il fotone l'opsina muta la sua conformazione, assume una serie di forme intermedie (*batorodopsina*) e dopo 1 ms è una *metarodopsina II*.

La figura 7.1 mostra il ciclo della rodopsina. La metarodopsina II rende accessibili i siti di legame specifici per vari regolatori, come la *rodopsin-chinasi* (*RK*), la *arrestina* (*antigene S* o *proteina 48-kDa*), la subunità α della transducina (*T α*), che partecipano all'amplificazione del segnale di *trigger*. La metarodopsina II lega alla membrana del disco una proteina globulare, la *transducina*,

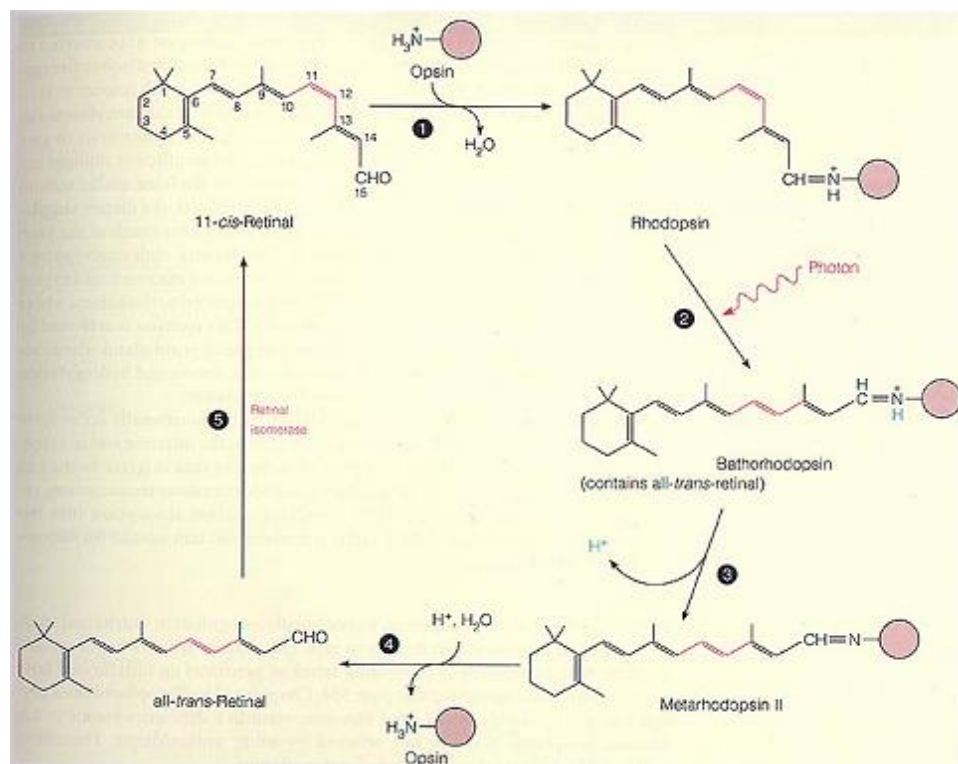


Figure 7: ciclo della rodopsina

un eterotrimerico composto da tre subunità ($T\alpha$, $T\beta$, $T\gamma$) differenziate per coni e bastoncelli. $T\alpha$ è una subunità catalitica, mentre $T\beta$ e $T\gamma$ sono subunità di regolazione. Nei bastoncelli le subunità sono $T\alpha 1$ (o GNAT1), $T\beta 1$ (o GNB1), e $T\gamma 1$ (o GNG1), mentre nei coni abbiamo $T\alpha 2$ (o GNAT2), $T\beta 3$ (o GNB3), e $T\gamma 8$ (o GNG8). La transducina è legata nel suo stato inattivo a una molecola di *guanosin-difosfato* (GDP). La metarodopsina II catalizza lo scambio di questo GDP con un *guanosin-trifosfato* (GTP); il gruppo $GTP-T\alpha$ si dissocia dal complesso membranario e si lega a una *fosfodiesterasi* (PDE), una proteina eterotetramerica attaccata alla membrana dei dischi, composta da quattro subunità (due catalitiche $PDE\alpha$, $PDE\beta$, e due inibitorie $PDE\gamma$); il gruppo $GTP-PDE\alpha\beta$ si dissocia dal complesso e può catalizzare l'idrolisi di un secondo messaggero, il $cGMP$ (*guanosin 3'-5' monofosfato ciclico*): $cGMP \rightarrow GMP + H^+$. La diminuzione di $[cGMP]$ determina la chiusura di parecchie centinaia di canali Na^+ normalmente aperti nell'oscurità, sicché la membrana esterna del bastoncello si iperpolarizza. Entro 1 sec dall'assorbimento di 1 fotone 400000 molecole di $cGMP$ sono state idrolizzate, con la conseguente chiusura di ≈ 300 canali Na^+ , pari al 3-5% di quelli aperti al buio. L'oscurità induce la formazione di nuova rodopsina e tende a riaprire i canali Na^+ depolarizzando la membrana esterna. Nell'oscurità è $V_m = -35mV$, Na^+ diffonde passivamente all'interno e viene estruso attivamente all'esterno per cui vi è una corrente I_{Na} nei due sensi (corrente di buio). La luce chiude i canali Na^+ e si ha iperpolarizzazione dovuta allo spostamento del potenziale di membrana verso il potenziale di equilibrio di altri ioni, la corrente di buio si riduce o cessa del tutto. Anche gli ioni Ca^{2+} , oltre agli ioni sodio, partecipano al processo di fototrasduzione, in un modo non del tutto chiarito.

La maggior parte dei dati pubblicati sulla trasduzione nei mammiferi si riferisce ai bastoncelli della retina dei bovini, o altri animali i cui campioni di tessuto sono facilmente reperibili. Nell'uomo esistono 4 proteine sensibili alla luce (fotopigmenti)

1. Rodopsina dei bastocelli (RHO). Assorb. max. a 495 nm
2. pigmento dei coni blu (*blue cone pigment*, BCP). Assorb. max. a 420 nm
3. pigmento dei coni verdi (*green cone pigment*, GCP). Assorb. max. a 530 nm
4. pigmento dei coni rossi (*red cone pigment*, RCP). Assorb. max. a 560 nm.

Tutti i fotopigmenti hanno la struttura a 7 alfa-eliche dei recettori associati alla proteina G. Sono inseriti nella membrana plasmatica dei cono e dei bastoncelli, e nei dischi membranosi dei bastoncelli. Un retinaldeide 11-cis è attaccato a un residuo lisina (RHO: Lys 296) nel settimo dominio transmembrana, mediante una base di Schiff.

La figura 7.1 mostra la sequenza primaria della rodopsina.



Figure 9: modello 3D della rodopsina

La rodopsina è un complesso che comprende

1. una regione extracellulare (E)
2. 7 α -eliche transmembrana
3. una superficie citoplasmatica (C)

La figura 7.1 mostra una ricostruzione tridimensionale.

La rodopsina è assai stabile. E' stato calcolato che l'isomerizzazione spontanea di una molecola di rodopsina per effetto termico dovrebbe verificarsi una volta ogni 3000 anni circa, cioè 1023 volte più lentamente rispetto alla fotoisomerizzazione (Baylor, 1987). L'energia richiesta per la fotoisomerizzazione del retinale è 45 kcal/mol. Tuttavia esiste una frazione ($< 0.01\%$) di molecole di retinale in cui il legame tra cromoforo e opsina è privo di protoni (Barlow e altri, 1993). Per queste molecole l'energia di fotoisomerizzazione scende a 23 kcal/mol, sicché possono contribuire al rumore di fondo al buio. Il granchio *Limulus* ha sviluppato un metodo per la riduzione del rumore di fondo dei fotorecettori retinici durante la notte, aumentando così la probabilità di localizzare un compagno. Pare che la percentuale di molecole di fotopigmento in stato privo di protoni venga ridotta per diminuzione del pH esterno [Barlow e altri, 1993].

Una mutazione nelle posizioni Pro23 o Gln28, nella regione extracellulare, provoca la *retinite pigmentosa*. Le eliche H1,4,6,7 sono ripiegate in corrispondenza dei residui Pro. Se la posizione di Gly90 è sostituita da un Asp si ha cecità al buio (per destabilizzazione del ponte salino). Il loop C-IV ha un certo numero di interazioni con peptidi che determinano l'attivazione della proteina G

7.2 Ribosomi

I *ribosomi* sono organelli presenti nel citoplasma dei procarioti e degli eucarioti, formati da proteine e un certo numero di molecole di rRNA prodotte nel nucleolo (3 nei procarioti, 4 negli eucarioti). Furono osservati per la prima volta al microscopio elettronico nel 1955 da G.E.Palade, sotto forma di granuli elettrodensi, attaccati alle membrane del reticolo endoplasmatico oppure liberi nel citoplasma.

I ribosomi sono particelle sferoidali con un diametro di circa 20-30 nm, formate da due subunità di dimensioni diverse, la *subunità grande* (*large subunit*, **LS**) e la *subunità piccola* (*small subunit*, **SS**), strettamente unite fra loro.

Il ribosoma è il luogo nel quale l'mRNA, copia complementare del DNA, è tradotto in proteine. Una volta che le due subunità LS e SS sono legate con l'mRNA proveniente dal nucleo, il ribosoma traduce l'mRNA in una sequenza specifica di aminoacidi, o catena polipeptidica.

Nelle cellule i ribosomi si presentano in due modi: *liberi* e *legati*. Essi si trovano in posizioni diverse a seconda della funzione della cellula. I *ribosomi liberi* vagano nel citosol, singolarmente o in gruppo; sono più numerosi di quelli legati in quelle cellule che trattengono la maggior parte delle proteine che producono; assemblano le proteine che vanno in soluzione nel citoplasma o che formano elementi mobili. In realtà i ribosomi liberi sono ancorati a fibre del citoscheletro. Uno stesso filamento di mRNA può essere tradotto da più ribosomi simultaneamente (*poliribosomi* o *polisomi*) come mostrato in figura 7.2.

I *ribosomi legati* sono attaccati sul lato esterno del reticolo endoplasmatico ruvido, sono più numerosi di quelli liberi nelle cellule che secernono le proteine che producono (cellule pancreatiche, produttori degli enzimi digestivi, ecc.); formano quelle proteine che divengono parte delle membrane o che sono immagazzinate in vescicole.

I ribosomi si trovano anche nei mitocondri e nei cloroplasti delle cellule eucariotiche, dove sintetizzano le proteine codificate dal DNA dei rispettivi organuli. Essi sono sempre più piccoli dei ribosomi citoplasmatici e sono comparabili ai ribosomi procariotici sia per la dimensione che per la sensibilità agli antibiotici. Ribosomi procariotici ed eucariotici eseguono le stesse funzioni attraverso le stesse reazioni chimiche.

Le cellule spendono molta energia per produrre ribosomi. Una cellula di *E. coli* può contenere 15000 ribosomi, ciascuno con peso molecolare $\approx 3 \times 10^6$ dalton, formando il 25% della massa totale di questo batterio. Le cellule dei mammiferi contengono $\approx 10^7$ ribosomi.

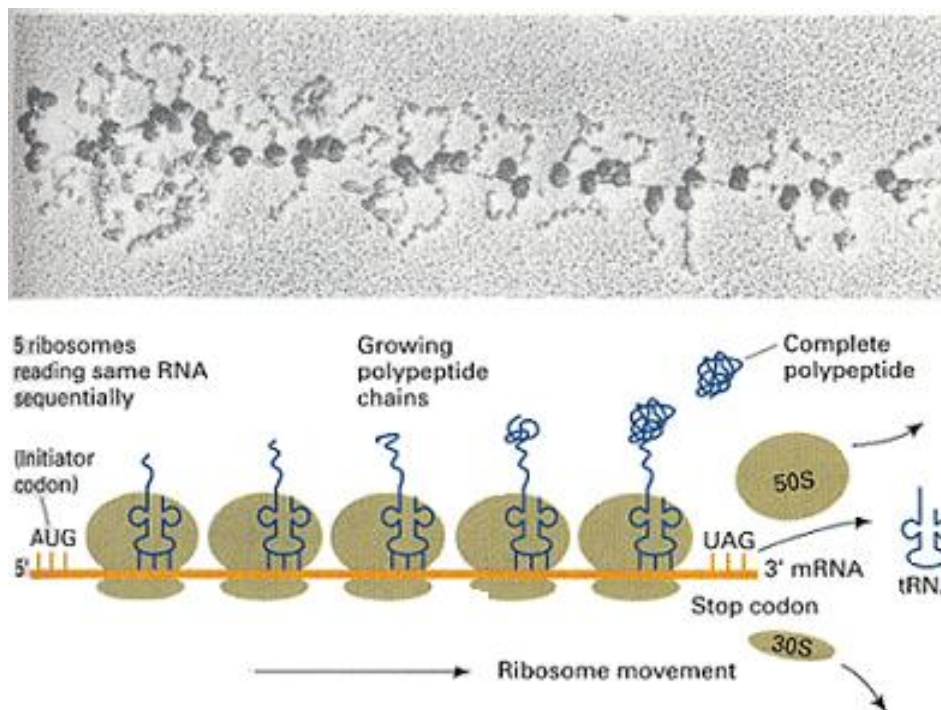


Figure 10: trascrizioni multiple simultanee

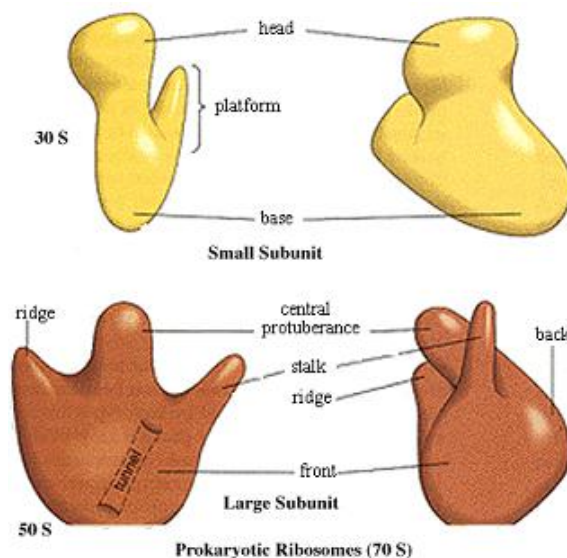


Figure 11: SS e LS del ribosoma procariotico

Definizione 7.2.1. Il valore di sedimentazione S di una macromolecola è, in unità [*Svedberg*], una misura della sua velocità di sedimentazione in una centrifuga, o del coefficiente di sedimentazione dopo centrifugazione in condizioni standard e in assenza di Mg^{2+} , ed è funzione sia del peso molecolare che della forma 3D.

Una unità Svedberg equivale a 10^{-13} sec. Il nome viene dal chimico svedese *Theodor Svedberg*, (1884-1971) inventore dell'ultracentrifuga, premio Nobel per la chimica nel 1926 per il suo lavoro sui sistemi dispersivi.

La forma dei ribosomi eucariotici e procariotici è molto simile. Le subunità hanno una *testa* e una *base*, con una *piattaforma* che protrude da un lato (vedi figura 7.2). La SS del ribosoma eucariotico ha pure un *becco* che dalla testa si estende verso il lato opposto della fessura, e un insieme di *lobi*. Si pensa che i lobi contengano le sequenze aggiuntive che rendono l'rRNA 18s più grande della controparte 16s del batterio.

La LS ha una protuberanza centrale detta *stelo*, e una *cresta* che si estende da un lato (vedi figura 7.2). Nella LS esiste un tunnel lungo circa 10 nm, con diametro 2.5 nm. Il tunnel si estende dalla regione contenente i siti A (*aminoacil*) e P (*peptidil*) alla zona da cui la nuova catena polipeptidica appena sintetizzata lascia il ribosoma. Si pensa che questo tunnel sia attraversato proprio dalla catena aminoacidica durante la sua formazione.

Una subunità è formata da molecole di RNA e proteine. Ogni subunità contiene almeno una molecola di rRNA. Le subunità tutte insieme contengono fino a 82 proteine specifiche. Le cariche positive delle proteine ribosomiali non

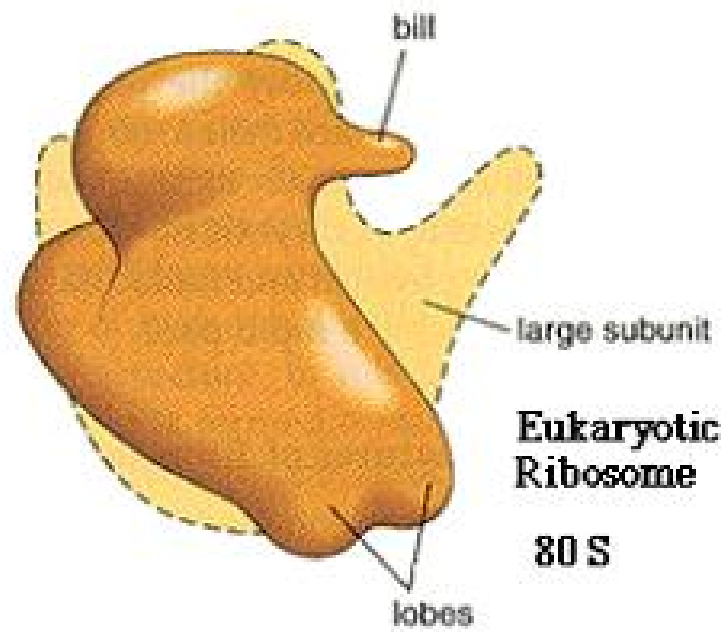


Figure 12: ribosoma eucariotico

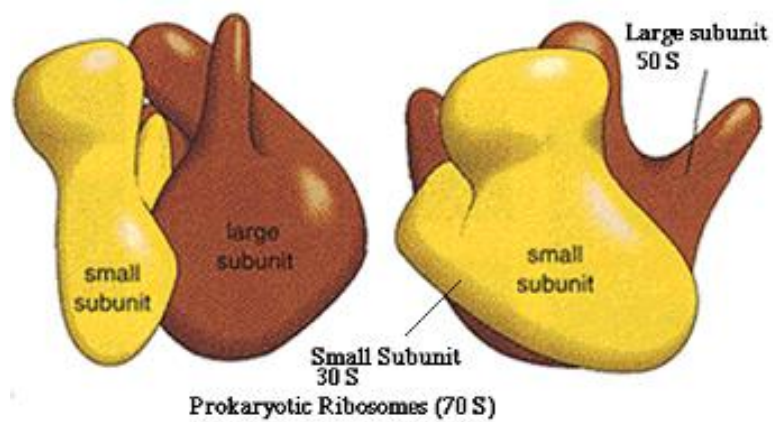


Figure 13: ribosoma procariotico

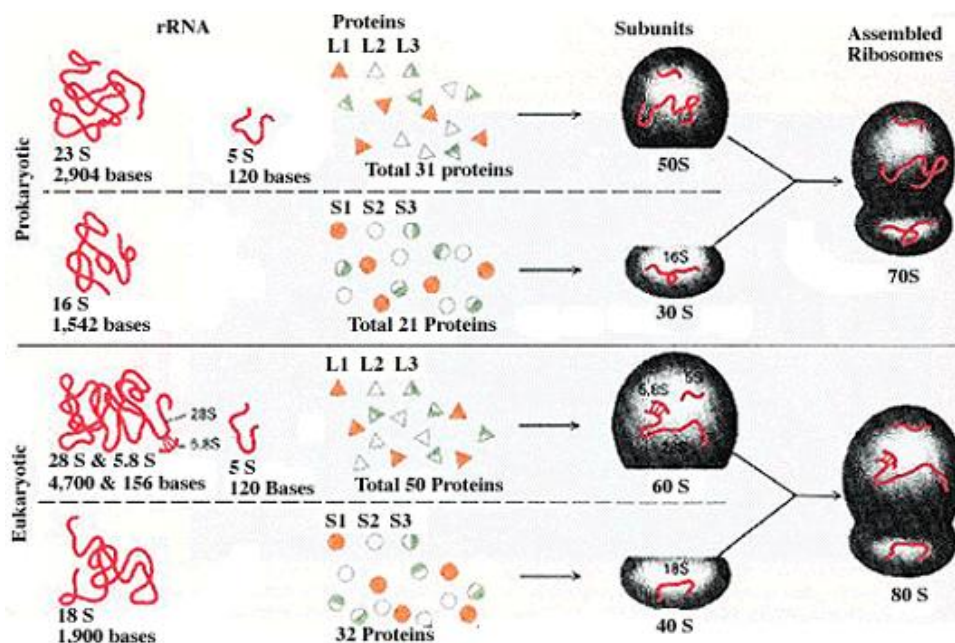


Figure 14: componenti ribosomiali

riescono a bilanciare i numerosi residui fosfato negativi dell'RNA, ciò fa sì che i ribosomi siano globalmente negativi, capaci quindi di legare cationi ed evidenziabili mediante coloranti basici.

La struttura delle proteine ribosomiali è stata ottenuta per lo più con la cristallografia a raggi X e con la NMR ad alta risoluzione.

Il *ribosoma procariotico* del batterio *E. coli* ha dimensioni 70s ed è formato da due subunità, di dimensioni 50s e 30s (vedi figura 7.2). Circa i 2/3 del ribosoma consistono di rRNA. La SS (30s) è formata da un rRNA (16s) e da 21 proteine ribosomiali, definite da S_1 a S_{21} in base alle dimensioni. Queste proteine sono unite tra loro da legami non covalenti. La LS (50s) ha 34 proteine ($L_1 - L_{34}$). Tutte queste proteine sono diverse tra loro, tranne L_{26} che è uguale a S_{20} , hanno p.m. compreso tra 7000 dalton e 32000 dalton, eccetto S_1 che pesa 65000 dalton.

I *ribosomi eucariotici* sono più complessi. Gli eucariotici semplici hanno i ribosomi più piccoli (anche se più grandi di quelli di *E. coli*) mentre i mammiferi hanno i ribosomi più grandi (80s), con subunità di dimensioni 60s e 40s. Il ribosoma contiene circa la stessa quantità in peso di rRNA e proteine.



Figure 15: sito di taglio di EcoRI

7.3 Enzimi di restrizione

Tagliando con il medesimo enzima di restrizione un segmento di cromosoma in un organismo qualunque e il DNA di un altro organismo e mescolando il tutto, è possibile che le relative estremità adesive si incontrino e si leghino dando origine ad una nuova sequenza ibrida di DNA. Per moltiplicare geni che interessano, il DNA che forma il gene d'interesse viene fatto ricombinare con il DNA appartenente a strutture biochimiche che hanno la capacità di passare da una cellula ad un'altra. Queste strutture possono essere i plasmidi, i batteriofagi o alcuni virus. Tra tutti vengono preferiti i plasmidi perché facilmente isolabili dal resto del DNA batterico e perché sono più maneggevoli degli altri potenziali vettori.

Definizione 7.3.1. Un *enzima di restrizione* (o *endonucleasi di restrizione*) è enzima capace di fissarsi su un sito particolare del DNA (*sito di restrizione*) e di realizzare un taglio dei due filamenti.

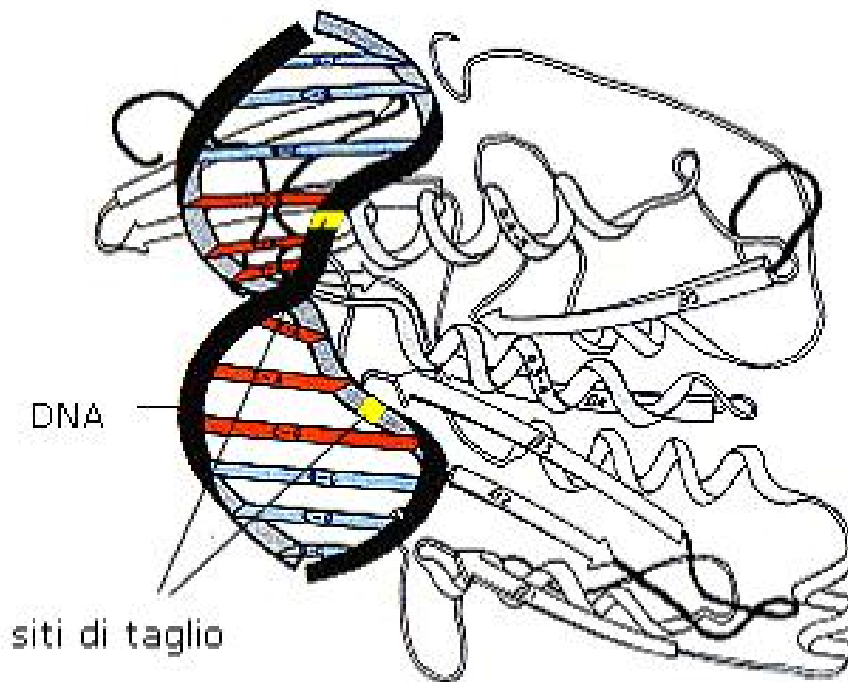
Le estremità del taglio avranno opportune sequenze, tra loro complementari, capaci di riappaiarsi (p.es. *TTAA* e *AATT*).

Gli enzimi di restrizione possono essere usati per tagliare segmenti di DNA da inserire nei plasmidi, anch'essi tagliati con lo stesso enzima di restrizione.

Definizione 7.3.2. Una *DNA-ligasi* è un'enzima catalizzatore della formazione di un legame tra due nucleotidi vicini su uno stesso filamento di DNA

In natura gli enzimi di restrizione sono prodotti da diverse specie batteriche, con la funzione difensiva di tagliare eventuali molecole di DNA estraneo che penetrino nella cellula batterica, p.es il DNA di un batteriofago. Per evitare che anche il proprio DNA venga attaccato dagli enzimi di restrizione, i batteri ospiti attaccano gruppi metilici ad alcune basi del proprio DNA, disattivando così l'azione enzimatica. Sono state finora identificate alcune centinaia di enzimi di restrizione diversi che eseguono il taglio di un qualsiasi DNA, aventi come siti di restrizione sequenze di 4-8 nucleotidi palindrome, cioè che si leggono allo stesso modo in una direzione, su un filamento, e nella direzione opposta, sull'altro filamento.

Inoltre, il DNA viene tagliato tra le stesse due basi su entrambi i filamenti: nel caso illustrato in figura, tra *G* e *A*. Si ottengono così due frammenti di DNA,



Una delle due subunita' dell'endonucleasi EcoRI

Figure 16:

uno con un'estremità di un filamento singolo TTAA e l'altro con un'estremità anch'essa di filamento singolo che si legge AATT. Le regioni complementari di DNA come queste (estremità adesive) possono appaiarsi e sono tenute unite da legami idrogeno tra le basi.

I ricercatori scegliendo di volta in volta, tra i vari enzimi di restrizione, possono utilizzare quelli che danno origine, tra gli altri frammenti di DNA, ad uno che contiene il gene che interessa. Mentre alcuni enzimi di restrizione tagliano nettamente entrambi i filamenti della molecola di DNA, altri li tagliano in modo che i due filamenti risultino leggermente sfalsati, cosicché le due estremità libere che si formano in seguito al taglio sono una un po' più lunga dell'altra. Per la tendenza ad appaiarsi che esiste tra le basi azotate complementari le due estremità "appiccicose" (sticky ends) possono aderire nuovamente alle estremità complementari di un'altra molecola di DNA che sia stata tagliata dallo stesso enzima di restrizione.

Gli enzimi di restrizione vengono denominati con sigle derivate dal nome del batterio che li contiene (ad esempio gli enzimi EcoRI dal batterio E.Coli). Essi permettono di tagliare in modo specifico la catena del DNA, in frammenti, detti frammenti di restrizione. Così anche frammenti di DNA di diversa origine possono unirsi purché prodotti con la stessa endonucleasi di restrizione. E' questa la chiave per la realizzazione di DNA ricombinante. I frammenti di due DNA d'origine diversa ottenuti con la stessa endonucleasi di restrizione, presentano le parti terminali che si appaiano e vengono saldate da un enzima ligasi. Un altro enzima, la transferasi terminale del timo, si usa per ottenere DNA ricombinante a partire da frammenti di restrizione con terminali pari, privi cioè di tratti sporgenti coesivi. Finora sono stati isolati molti enzimi di restrizione da diverse specie di batteri. La specificità e la varietà degli enzimi di restrizione hanno permesso ai genetisti molecolari di identificare e isolare segmenti del DNA di molti organismi, compreso l'uomo.

8 Appendice II. Dinamica molecolare

8.1 Introduzione

Uno degli strumenti principali nello studio delle molecole biologiche è quello delle *simulazioni di Dinamica Molecolare (DM)*. E' un'applicazione di tecniche computazionali per determinare l'evoluzione temporale di un sistema molecolare. Tali simulazioni forniscono preziose informazioni sulle vibrazioni locali e sulle transizioni conformazionali delle proteine e degli acidi nucleici. Gli obiettivi primari sono la determinazione della struttura, della dinamica e termodinamica delle molecole biologiche e dei loro complessi. Dati complementari vengono dalla cristallografia a raggi X e dagli esperimenti NMR.

Le simulazioni ricostruiscono le traiettorie dei nuclei atomici con una risoluzione di $10^{-15} - 10^{-15}$ sec, coprendo un arco di tempo tipicamente dell'ordine di 10^{-12} sec. Tre tipi di transizioni energetiche si verificano comunemente nelle molecole:

1. rotazionali, anche a temperatura ambiente ($E = 0.4 - 2 \times 10^{-23}$ J)
2. vibrazionali, ($E = 2 - 20 \times 10^{-21}$ J)
3. elettroniche, ($E = 2 \times 10^{-19}$ J)

Se le transizioni non comportano modificazioni drastiche nella struttura elettronica, le energie possono essere calcolate usando la meccanica molecolare classica (programmi come AMBER, CHARMM, Discover, QUANTA/CHARMM, SYBYL), se invece vi sono modificazioni chimiche (rottura o formazione di nuovi legami) si ricorre ai metodi della meccanica quantistica (programmi come CHARMM/GAMESS).

I processi caratteristici che avvengono nelle molecole biologiche si svolgono su differenti scale spazio-temporali:

1. movimenti locali ($0.01-5 \text{ \AA}$, $10^{-15} - 10^{-1} s$)
 - (a) fluttuazioni atomiche
 - (b) movimenti sidechain
 - (c) movimenti ciclici
2. movimenti rigidi ($1-10 \text{ \AA}$, $10^{-9} - 1 s$)
 - (a) movimenti di elica
 - (b) movimento di domini (flessione in punti di snodo)
 - (c) movimenti di subunità
3. movimenti su larga scala ($>5 \text{ \AA}$, $da 10^{-7} - 10^4 s$)
 - (a) transizioni di anello delle elica
 - (b) Dissociazione/Associazione

(c) denaturazione

In linea di principio si potrebbe partire dall'equazione di Schroedinger dipendente dal tempo per tutti i nuclei e gli elettroni della molecola

$$\frac{\hbar}{2m} \nabla^2 \Psi(\mathbf{r}, t) + V(\mathbf{r}, t) \Psi(\mathbf{r}, t) = \frac{\hbar}{i} \frac{\delta \Psi(\mathbf{r}, t)}{\delta t} \quad (8.1.1)$$

dove l'energia potenziale $V(\mathbf{r}, t)$ è data dalle forze coulombiane di attrazione/repulsione che nuclei ed elettroni esercitano su una determinata particella del sistema. La *funzione d'onda* $\Psi(\mathbf{r}, t)$ è soluzione della (8.1.1) e fornisce la probabilità di trovare la particella (elettrone o nucleo) in una data posizione \mathbf{r} e in un dato istante t .

La funzione d'onda può essere espressa come prodotto di una funzione indipendente dal tempo e di una funzione dipendente solo dal tempo

$$\Psi(\mathbf{r}, t) \equiv \Psi(\mathbf{r}) f(t) \quad (8.1.2)$$

Riscrivendo la (8.1.1) abbiamo

$$\frac{\hbar}{2m} \nabla^2 \Psi(\mathbf{r}) + V(\mathbf{r}) \Psi(\mathbf{r}) = E \Psi(\mathbf{r}) \quad (8.1.3)$$

dove E è l'energia e $V(\mathbf{r})$ l'energia potenziale. La (8.1.3) è talvolta scritta in notazione abbreviata

$$H\Psi = E\Psi \quad (8.1.4)$$

In pratica ci sono diverse ragioni per le quali non è conveniente seguire questo approccio rigoroso:

1. soluzioni analitiche esistono solo in un certo numero di casi semplici (atomo H, oscillatore armonico, particella in un contenitore, ecc.). Per sistemi con molte particelle non si è giunti a nulla di analogo in 60 anni di ricerca
2. spesso le soluzioni approssimate della (8.1.1) funzionano bene e talvolta permettono una comprensione maggiore dei fenomeni (calcolo di Huckel, ecc.).

Queste approssimazioni sono la base delle simulazioni di dinamica molecolare.

8.2 Approssimazione di Born-Oppenheimer

L'approssimazione di Born-Oppenheimer permette di separare nella funzione d'onda il contributo dei nuclei da quello degli elettroni

$$\Psi(\mathbf{r}) \equiv \Psi(\mathbf{r})^{el} \Psi(\mathbf{r})^n \quad (8.2.1)$$

Questa operazione non è formalmente corretta, ma è giustificata dal fatto che la massa dell'elettrone è molto minore di quella di un qualsiasi nucleo. L'importanza dell'approssimazione viene dal fatto che grazie ad essa si introduce il concetto di *superficie di energia potenziale*.

Ringraziamenti

I miei ringraziamenti a...

References

- [1] Schneider, T.D. *Theory of molecular machines, I. Channel capacity of molecular machines*, J. Theor. Biol. 148, 83-123.
- [2] Schneider, T.D. *Theory of molecular machines, II. Energy dissipation from molecular machines*, J. Theor. Biol. 148, 125-137.
- [3] Schneider, T.D. *Sequences Logos, Machine/Channel Capacity, Maxwell's Demon, and Molecular Computers: a Review of the Theory of Molecular Machines*, Nanotechnology, 5(1):1-18, 1994.
- [4] Schneider, T.D. *Measuring Molecular Information*, Journal of Theoretical Biology, 201: 87-92, 1999.
- [5] Schneider, T.D. *Evolution of biological information*, Nucleic Acids Research, Vol.28, No.14: 2794-2799, 2000.
- [6] Schneider, T.D. *New Approaches in Mathematical Biology: Information Theory and Molecular Machines*, in Julian Chela-Flores and Francois Raulin, *Chemical Evolution: Physics of the Origin and Evolution of Life*: 313-321, Dordrecht, Netherlands, 1996. Kluwer Academic Publishers
- [7] Schneider, T.D. *Some Lessons for Molecular Biology from Information Theory*, in *Entropy Measures, Maximum Entropy and Emerging Applications*, Special Series on *Studies in Fuzziness and Soft Computing*, Vol. 19: 229-237, NY, 2003. Springer-Verlag.
- [8] Schneider, T.D. *Information Content of Individual Genetic Sequences*, Journal of Theoretical Biology, 189 (4): 427-441, 1997.
- [9] siti vari sul WWW
- [10] K. Palczewski, T. Kumasaka, T. Hori, C.A. Behnke, H.Motoshima, B.A. Fox, I. Le