

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

---

Scuola di Scienze  
Dipartimento di Fisica e Astronomia  
Corso di Laurea in Fisica

## Sequenziamento del DNA tramite nanopori

**Relatore:**

**Prof. Gastone Castellani**

**Presentata da:**

**Pietro Miguel Bertoldi**

Anno Accademico 2015/2016

*Alla Mondeo ed il suo equipaggio,  
a chi ha preso il largo e a chi è rimasto in porto,  
ché con prua sicura e coraggio,  
han trovato quel tesoro che è un sol pezzo da otto!  
Nessun oceano nè alpe li potrà mai più separare,  
ché ognuno ha la sua parte!  
Abbate vento in poppa e maledite la vita in bonaccia!*

## Sommario

L'elaborato illustra alcuni metodi di sequenziamento del DNA, partendo da dei cenni sulle tecniche a sfondo chimico tradizionali e di nuova generazione più diffuse.

Tratta il sequenziamento tramite nanopori, descrivendo i fenomeni fisici nanoscopici su cui si basa, alla luce delle caratteristiche e delle proprietà degli acidi nucleici e dei pori in questione, concentrandosi in particolare sul nanoporo biologico  $\alpha$ -emolisina.

Di questa tecnica mostra i risultati raggiunti, le potenzialità ancora da sviluppare e i vantaggi finora emersi rispetto alle altre in uso da più tempo.

# Indice

<b>Introduzione</b>	<b>3</b>
<b>1 Il DNA e le sue caratteristiche fisiche</b>	<b>4</b>
1.1 Metilazione del DNA . . . . .	10
<b>2 Sequenziamento del DNA</b>	<b>11</b>
2.1 Metodo di Sanger . . . . .	11
2.2 Metodi di nuova generazione . . . . .	14
2.2.1 Pirosequenziamento . . . . .	14
2.2.2 Sequenziamento a colorante . . . . .	15
<b>3 Sequenziamento tramite nanopori</b>	<b>17</b>
3.1 Traslocazione del DNA . . . . .	18
3.2 Nanopori sintetici . . . . .	21
3.3 Nanopori Alpha emolisina . . . . .	23
3.3.1 Caratteristiche del nanoporo . . . . .	23
3.3.2 Corrente ionica di ostruzione . . . . .	24
3.3.3 Corrente trasversa . . . . .	26
3.3.4 Siti di rivelazione . . . . .	30
3.3.5 Risultati del sequenziamento tramite Alpha emolisina . . . . .	32
3.4 Nanopori Mycobacterium smegmatis porin A . . . . .	33
3.4.1 Caratteristiche del nanoporo . . . . .	33
3.4.2 Corrente ionica e siti di rivelazione . . . . .	33

3.4.3 Risultati del sequenziamento tramite <i>Mycobacterium smegmatis</i> porin A . . . . .	38
<b>4 Conclusioni</b>	<b>40</b>
<b>Bibliografia</b>	<b>41</b>

# Introduzione

Il sequenziamento degli acidi nucleici, DNA, RNA e loro declinazioni in genere, è il processo che ne rivela la sequenza nucleotidica, cioè l'ordine delle basi azotate. Queste sequenze codificano le informazioni genetiche ereditarie di un organismo, è quindi evidente l'importanza di sviluppare tecniche, veloci, affidabili e a basso costo, che permettano di leggerle. Questa capacità consentirebbe di ottenere informazioni sul corredo genetico delle diverse specie viventi, specie estinte, sui meccanismi che regolano l'espressione genica, sulle cause e gli effetti delle mutazioni, con grande velocità, efficienza e pochi costi.

I campi in cui può essere applicata sono svariati, biologia, astro-biologia, medicina etc.. La medicina "personalizzata", basata, cioè, sul genoma individuale del paziente, è oggi un campo all'avanguardia della ricerca scientifica. Questo è stato reso possibile grazie al miglioramento dei metodi di sequenziamento del DNA. Per mettere in pratica il suo potenziale, è necessario un metodo innovativo, che sia veloce e a basso costo. Questa possibilità può essere fornita da un approccio fisico al problema, meno limitato dalla lunghezza dei campioni e dal tempo di analisi (principali limiti dei metodi chimici attuali). La possibilità di sfruttare determinate caratteristiche fisiche del DNA, per poterlo sequenziare, è stata fornita dallo sviluppo di tecnologie che implementano nanopori, sintetici e biologici.

Inoltre, il percorso per arrivare ai metodi che impiegano nanopori e sonde nanometriche, pone nuove sfide alla nostra comprensione, capacità di manipolare e studiare processi fisici, che coinvolgono solidi, liquidi e biomolecole, alla scala nanometrica.

# Capitolo 1

## Il DNA e le sue caratteristiche fisiche

Dal momento che quest'elaborato tratta di sequenziamento del DNA, si ritiene opportuno cominciare con una descrizione degli acidi nucleici e dei nucleotidi che li compongono, da un punto di vista biologico e fisico, concentrando l'attenzione su quelle caratteristiche fondamentali ai fini del processo in questione.

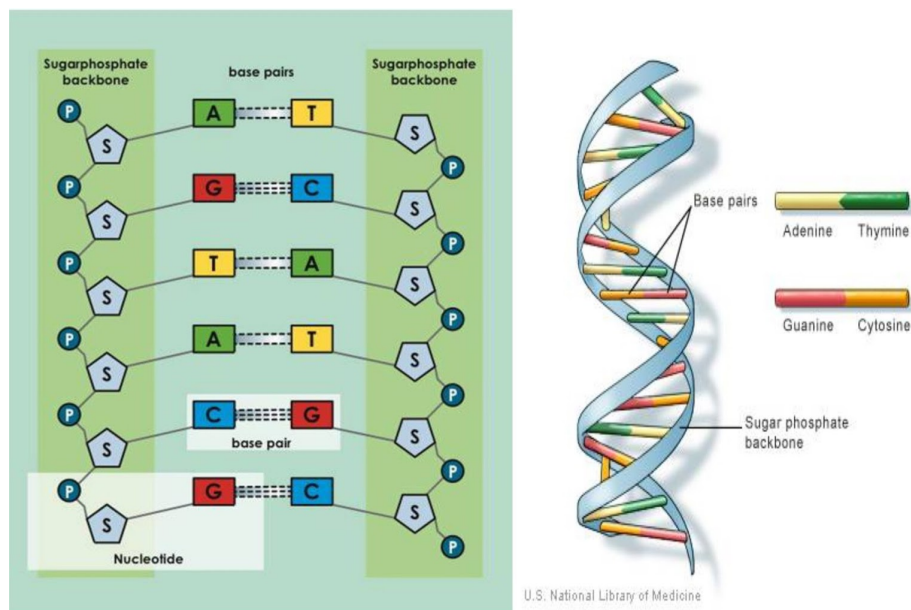


Figura 1.1: Schematizzazione della struttura del DNA. A sinistra le impalcature di zuccheri (S) e gruppi fosfato (P) con le paia di basi, al centro la struttura secondaria a doppia elica del ds-DNA, a destra le coppie di basi possibili A+T, G+C

Un nucleotide, è una biomolecola formata da tre componenti caratteristici: una base azotata, uno pentosio (zucchero pentoso) e un gruppo fosfato. Le basi azotate che ci interessano sono cinque, suddivise in purine e pirimidine. Le prime sono adenina (A) e guanina (G), le seconde citosina (C), timina (T) ed uracile (U).

Un acido nucleico è un polimero composto da una successione di nucleotidi. Il pentosio si lega al gruppo fosfato dell'unità successiva, creando un'impalcatura (scheletro), che segue la direzione  $3' \leftrightarrow 5'$  dei pentosi, sulla quale sono legate le basi azotate come fossero gruppi laterali.

Gli acidi nucleici più conosciuti sono il DNA e l'RNA, che si distinguono per il pentosio che contengono, rispettivamente desossiribosio e ribosio. Il primo contiene sequenze di adenina, guanina, citosina e timina, mentre il secondo sostituisce la timina con l'uracile. Questa caratteristica però non li distingue in quanto raramente si incontrano eccezioni.

## Struttura

Grazie al noto modello costruito da Watson e Crick, con il fondamentale contributo di Rosalind Franklin, nel 1953, che, pur essendo molto intuitivo, riproduce abbastanza fedelmente quella che oggi si può affermare essere la struttura del DNA, si sa che questa molecola è costituita da due filamenti antiparalleli di nucleotidi avvolti l'uno sull'altro a formare una doppia elica. Le basi azotate dei nucleotidi si legano a coppie, tramite legami a idrogeno, unendo i due filamenti, rimanendo all'interno della doppia elica come i pioli di una scala, mentre gli zuccheri e i gruppi fosfati uniscono i nucleotidi l'uno all'altro a formare i singoli filamenti, costituendone l'impalcatura esterna. Le basi azotate si legano solo in coppie definite: C+G o A+T/U. Questo fa sì che i due filamenti che compongono la doppia elica siano complementari, quindi ottenutone uno, si può risalire alla composizione dell'altro.

Esiste una proprietà chiamata "struttura secondaria". Per il ds-DNA (doppio filamento), ad esempio, esistono vari tipi di eliche. Le più comuni sono A-DNA e B-DNA, ciò che cambia la struttura d'insieme fra A e B sono le condizioni ambientali, ioniche e della soluzione. Il B-DNA è più frequente in soluzioni acquose in quanto le molecole d'acqua si legano nelle scanalature lungo l'elica. Questa struttura è caratterizzata da un distanza fra le paia di basi di 3.4 e un angolo fra loro di  $36^\circ$ , che danno 10 paia di basi



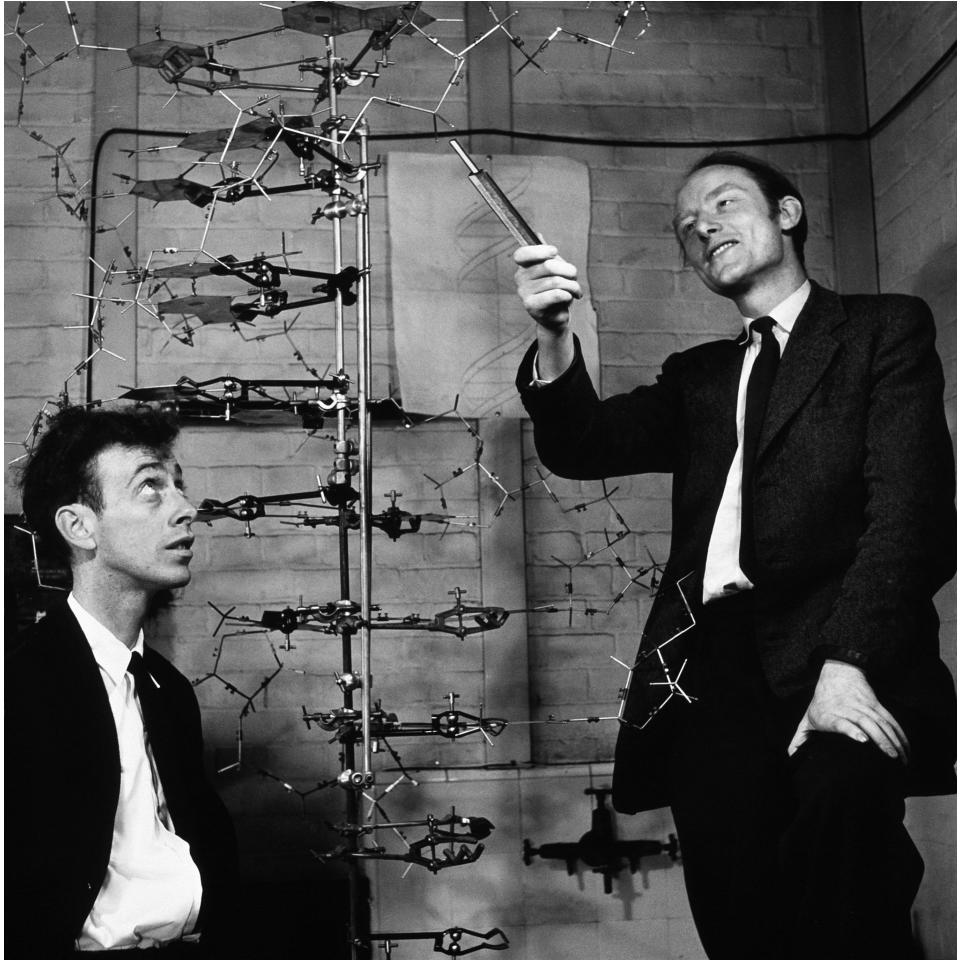


Figura 1.2: James Watson e Francis Crick mostrano il loro modello a doppia elica del DNA

per avvolgimento e un diametro dell'elica di 2 nm. Queste informazioni potrebbero non essere fondamentali per il sequenziamento, ma sono utili per altri scopi di rivelazione del ds-DNA.

Anche singoli filamenti di DNA (ss-DNA) possono mostrare una struttura secondaria dipendente dalle condizioni ioniche e dalla temperatura. In generale la struttura secondaria in questo caso è determinata da fattori entalpici ed entropici. Quando prevalgono i primi avremo un'elica stretta e ordinata, mentre se predominano i secondi, il filamento potrà assumere varie configurazioni casuali, dovute a rotazioni dei gradi di libertà dell'impalcatura del filamento.

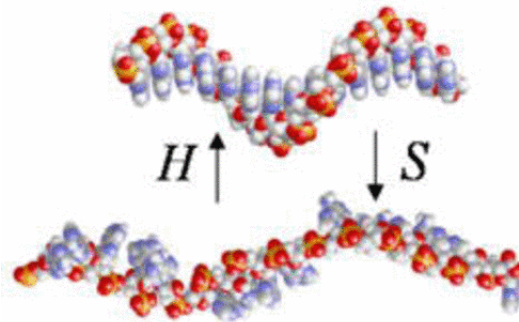


Figura 1.3: Schematizzazione degli effetti entalpici (H) ed entropici (S) sulla struttura secondaria di un singolo filamento di DNA

Fra le sequenze particolari di basi azotate nel DNA, una abbastanza comune è il palindromo, cioè una sequenza che si legge allo stesso modo sia da sinistra verso destra, sia da destra verso sinistra. Questo termine si riferisce a regioni del DNA con ripetuti invertiti di una sequenza di basi. Queste regioni sono dunque autocomplementari, perciò possono formare strutture secondarie dette "a forcina" o "a croce". In particolare ci interessano le prime, in quanto riguardano anche i filamenti singoli, e possono creare problemi durante la traslocazione del DNA attraverso un nanoporo nel processo di sequenziamento.

Una particolare tecnica per poter trascurare le strutture secondarie, è quella di usare un'enzima, l'exonuclease, che stacca uno a uno i nucleotidi dal filamento. In questo modo si possono analizzare singolarmente uno per volta.

Tabella 1.1: Volumi delle singole basi e dei rispettivi nucleotidi  $V_B$  e  $V_N$  espressi in  $\text{\AA}^3$ , fra parentesi le aree delle sezioni  $A_B$  e  $A_N$  espresse in  $\text{\AA}^2$ . L'ultima colonna riporta la frazione di volume libero, definita nel paragrafo 3.3.2, con nanopori di diametro 15 e 20  $\text{\AA}$

Base	$V_B(A_B)$	$V_N(A_N)$	$F_{15}(F_{20})$
A	157 (166)	349 (340)	0.72 (0.841)
G	168 (177)	359 (351)	0.71 (0.837)
C	133 (147)	324 (319)	0.74 (0.853)
T	150 (163)	339 (331)	0.73 (0.846)

A questo punto potremmo calcolare le dimensioni di ogni nucleotide a seconda della base azotata che li compone, ma come possiamo notare dai dati, quest'informazione da

sola non è sufficiente alla discriminazione.

## Elettronica

Alcuni metodi di rivelazione e sequenziamento, ipotizzano di sfruttare variazioni di tensione e trasporto elettronico per discriminare le basi. Queste tecniche sono sensibili alla struttura elettronica dei nucleotidi, alla loro geometria e all'ambiente circostante.

Il trasporto attraverso i nucleotidi, rivela differenze negli stati elettronici tramite la loro energia ed estensione spaziale. Per conoscere le differenze fra le basi, vengono calcolate le densità di stati per ognuna e per l'impalcatura, ponendole fra elettrodi d'oro. Dal grafico in figura (1.4), si nota che i nucleotidi non hanno una struttura elettronica molto diversa, avendo stati molecolari energeticamente molto vicini comparati al livello di Fermi degli elettrodi. In particolare il contributo maggiore di stati al livello di Fermi, è dato dall'impalcatura, in quanto ha la superficie di contatto maggiore con un elettrodo. Le basi, in prossimità del livello di Fermi contribuiscono in maniera diversa alla densità di stati a seconda della loro geometria, estensione e orientazione spaziale.

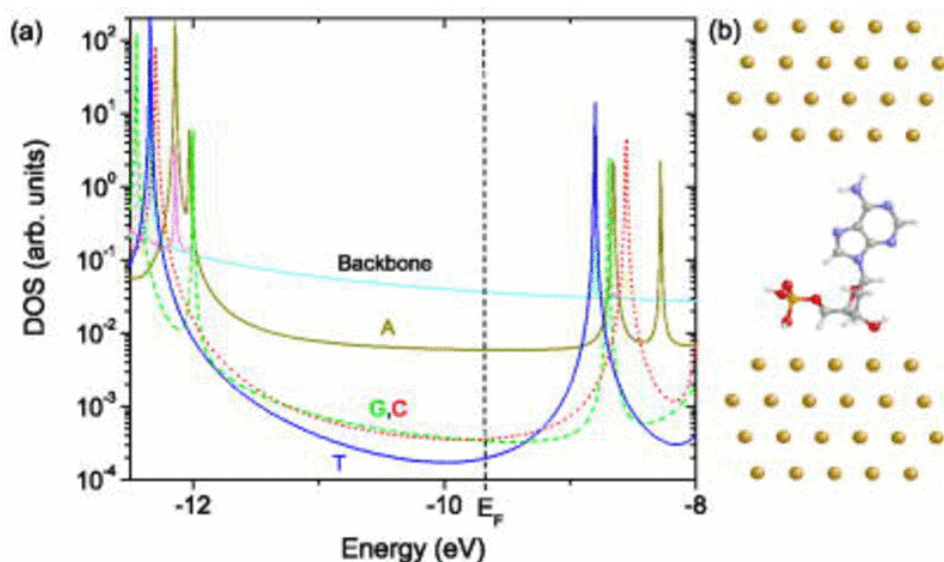


Figura 1.4: Le densità di stati delle singole basi, dell'impalcatura di zuccheri e fosfati, il livello di fermi è riferito agli elettrodi d'oro

Tuttavia, interessa sapere che il contributo base, all'accoppiamento nucleotide-elettrodo, è determinato dalla natura degli stati molecolari. Mostriamo in figura (1.5) gli schemi per basi isolate, secondo il più alto orbitale molecolare occupato (HOMO) e per il più basso orbitale molecolare non occupato (LUMO). Saranno questi a determinare in che modo un nucleotide poco a contatto con l'elettrodo si accoppierà ai portatori di carica di quest'ultimo. Notiamo che, in generale, per tutte le basi gli stati sono distribuiti attorno agli anelli, e rimarranno così anche in presenza dell'impalcatura carica (potrebbero però non essere più HOMO e LUMO).

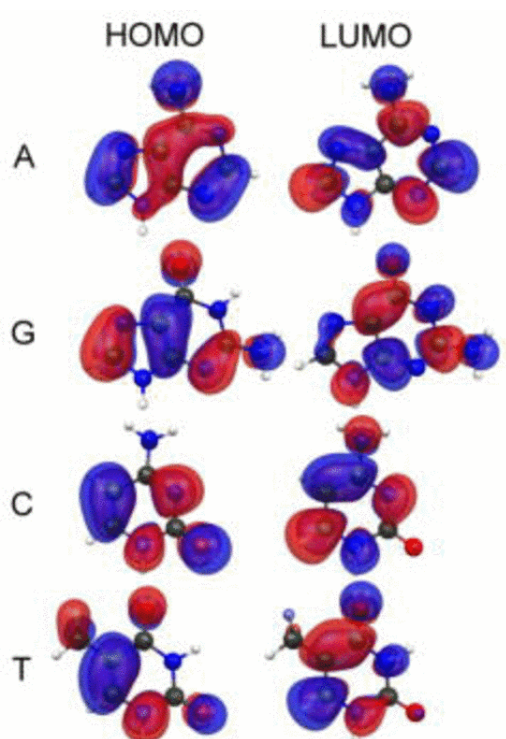


Figura 1.5: Iso-superfici degli stati HOMO e LUMO delle basi azotate

Un altro approccio elettronico è quello di misura il momento di dipolo di ogni base, le cui intensità sono mostrate in tabella sia per le basi libere che per i corrispondenti nucleotidi. Sia i momenti di dipolo, che le configurazioni elettroniche, sono state calcolate utilizzando l'approssimazione di Hartree-Fock.

L'ultima proprietà elettronica importante, ai fini del sequenziamento, è il modo in cui i polinucleotidi interagiscono e si legano alle le superfici, in particolare quelle interne

Tabella 1.2: Momenti di dipolo elettrico molecolari, espressi in D (Debye), calcolati teoricamente in approssimazione di Hartee-Fock per le basi  $p_B$  ed i nucleotidi  $p_N$ , confrontati con i valori ottenuti sperimentalmente  $p_B^{expt}$

Base	$p_B$	$p_B^{expt}$	$p_N$
A	2.33	2.5	4.76
G	7.17	7.1	7.76
C	7.22	7.0	8.55
T	4.72	4.1	7.56

dei nanopori. Di ciò si rimanderà la trattazione in dettaglio più avanti.

## 1.1 Metilazione del DNA

La metilazione è un processo durante il quale un gruppo metile ( $-\text{CH}_3$ ) si lega ad una molecola, nel DNA in particolare, si lega ad una base azotata. La metilazione del DNA ne è dunque una modifica epigenetica. Tutti i diversi nucleotidi possono esserne oggetto, assumendo diverse funzioni. Particolarmente interessante è la metilazione della citosina, che è il tipo più presente negli organismi eucarioti e ne regola l'espressione genica, promuovendo o inibendo determinati geni. La metilazione del DNA è molto importante, in quanto regola processi come l'imprinting genomico, l'inibizione del cromosoma X; la sua assenza, poi, può portare alla carcinogenesi. È importante quindi, che un metodo di sequenziamento sia in grado di discriminare la dC dalla mC. Permettendo di acquisire ulteriori conoscenze sulle funzioni di questo processo, sia in campo medico (cancro, malattie genetiche), che biologico.

# Capitolo 2

## Sequenziamento del DNA

Considerato che il DNA è il "database" in cui è contenuta l'informazione genetica di un organismo, è evidente che la sua più importante proprietà è la sequenza nucleotidica, che codifica i dati come il codice binario nei dispositivi elettronici. La capacità di poter leggere e tradurre questo codice è dunque di cruciale importanza in svariati campi, medicina, biologia, biotecnologia etc..

La possibilità di determinare sequenze di nucleotidi arrivò verso la fine degli anni '70, grazie allo sviluppo della conoscenza della chimica dei nucleotidi, del metabolismo del DNA e di tecniche avanzate per l'elettroforesi. Si delinearono due metodi, Maxam-Gilbert e il metodo di Sanger, descriveremo solamente il secondo, in quanto è stato il più usato (viene usato tutt'ora) e concettualmente è simile al primo. In entrambe le tecniche infatti, il principio generale si basa sulla riduzione del DNA da sequenziare a quattro gruppi di frammenti marcati. La reazione che produce questa suddivisione è sensibile al tipo di base, dunque si ottengono quattro campioni di frammenti di diverse lunghezze che terminano con una specifica base per gruppo.

### 2.1 Metodo di Sanger

Questo metodo utilizza la sintesi di DNA per mezzo di una DNA-polimerasi, un enzima impiegato nelle cellule per la duplicazione del DNA. La DNA-polimerasi necessita di un primer (una catena oligonucleotidica) a cui aggiungere i nucleotidi, e un filamento stam-

po che ne determini la scelta. Quindi questo enzima, utilizzando il ssDNA che vogliamo sequenziare come stampo, creerà nuovi filamenti complementari, che verranno poi impiegati nel processo vero e proprio, conservando in questo modo il filamento originale. Il primer interagisce con deossinucleotidi trifosfati (dNTP), a cui si lega per prolungarsi, se incontrasse un dideossinucleotide, non sarebbe in grado di formare il legame successivo, interrompendo dunque la sintesi.

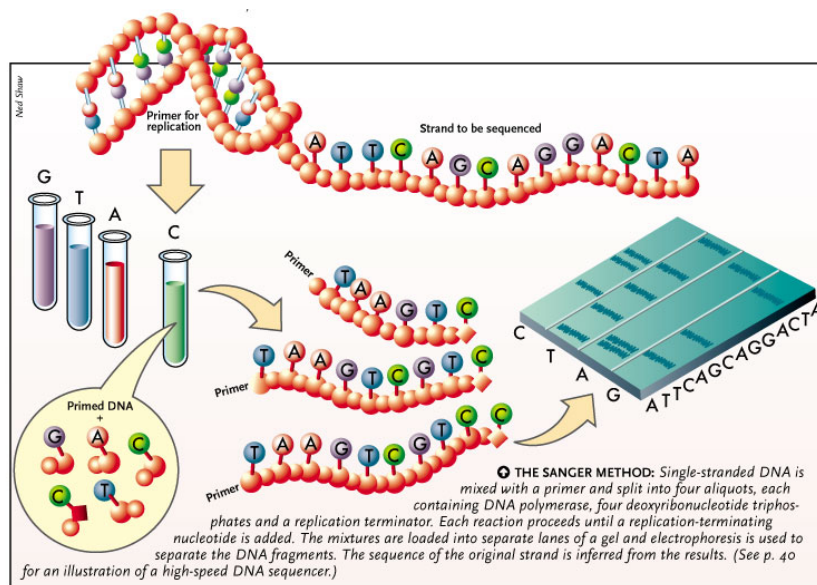


Figura 2.1: Illustrazione del metodo di Sanger per corsa elettroforetica

Il DNA da sequenziare, usato come filamento stampo, i primer utilizzati contengono radioisotopi o vengono resi fluorescenti. Si inseriscono in soluzione anche ddNTP e dNTP, i secondi in grande eccesso rispetto ai primi. Questo avviene per quattro campioni uguali, uno per base. Ogniqualvolta la DNA-polimerasi incontra un ddNTP, interrompe la sintesi, creando un filamento di una specifica lunghezza che termina con una specifica base. Dato che i dNTP sono in eccesso rispetto ai ddNTP (che comunque è in una quantità sufficiente), avremo tutti i frammenti corrispondenti alle diverse posizioni della base N nel campione.

A questo punto abbiamo ottenuto quattro insiemi divisi per base, contenenti frammenti di diverse lunghezze, che possiamo separare mediante elettroforesi, i più corti arriveranno più lontano, i più lunghi, più lenti nel gel, percorreranno una distanza più

breve. Sapendo che ad ogni pista corrisponde una base, si può determinare la sequenza del campione in esame (ricordando che abbiamo studiato elettroforeticamente il suo complementare).

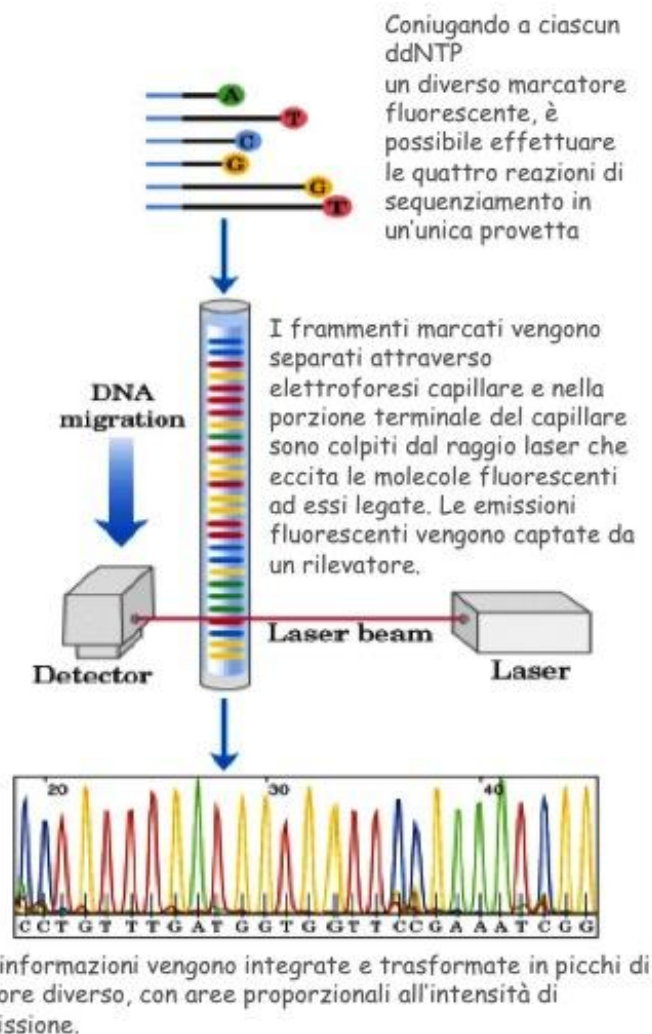


Figura 2.2: Illustrazione del metodo di Sanger automatizzato

Il radioisotopo è necessario per poter osservare l' autoradiografia del gel, ma se al suo posto venisse impiegata una molecola fluorescente, potremmo automatizzare e velocizzare il processo. Assegnando ad ogni base una molecola fluorescente di un colore specifico, e facendo correre i frammenti sulla stessa pista nel gel, con un laser e un rivelatore



saremmo in grado di distinguere ogni base al momento del suo passaggio attraverso il laser (nel caso in cui ci siano più frammenti allo stesso istante, avremo una distribuzione del segnale molto piccata, che non costituisce alcun problema).

Questo metodo, nella sua semplicità concettuale, è molto laborioso nella realizzazione. Sequenziare polinucleotidi molto lunghi richiede tempi e costi alti, dettati dai numerosi reagenti per la preparazione dei campioni e dai tempi lunghi intrinseci dell'elettroforesi per gruppi numerosi di filamenti.

## 2.2 Metodi di nuova generazione

I metodi di nuova generazione, o ad alto parallelismo, sono la risposta alla richiesta di tecnologie per un sequenziamento rapido e poco costoso. Questi metodi, parallelizzano il processo di sequenziamento producendo un grande numero di sequenze contemporaneamente. Ne citeremo e descriveremo brevemente solamente alcuni, per poi analizzare più in dettaglio in sequenziamento tramite nanopori, più interessante dal punto di vista fisico.

### 2.2.1 Pirosequenziamento

Il pirosequenziamento è una tecnica basata sul sequenziamento per sintesi. Si distingue dal metodo di Sanger in quanto non sfrutta la fluorescenza e l'elettroforesi per determinare la sequenza. Misura il rilascio di pirofosfato inorganico, che viene convertito proporzionalmente in luce visibile da una serie di reazioni enzimatiche. Diversamente da altre tecniche che usano dNTP modificati per terminare la sintesi del DNA, il pirosequenziamento gestisce la DNA-polimerasi con singole introduzioni di dNTP in quantità limitate. In seguito all'aggiunta dei nucleotidi complementari, la sintesi prosegue estendendo il filamento, per poi arrestarsi quando incontra una base non complementare. Il campione viene quindi pulito dalle basi in eccesso e ne vengono inserite altre di diverso tipo. Questo procedimento è automatizzato e ripetuto ciclicamente in modo rapido, la luce generata dalle reazioni enzimatiche viene acquisita tramite una camera CCD, ed associata alla base presente nel sistema in quell'istante. Tutti i segnali luminosi ven-

gono quindi graficati (intensità maggiori corrispondono a sequenze di più basi uguali) nel "pirogramma", dal quale si evince la sequenza complementare del DNA.

Per quanto semplice nel concetto e nella realizzazione, il pirosequenziamento ha alcune limitazioni. La più importante è che la lunghezza delle sequenze è solitamente inferiore alle 100 basi, che ne fa un metodo molto utile per quanto riguarda l'identificazione di brevi sequenze come per la SAGE (Analisi Seriale dell'Espressione Genica) o SNP (Polimorfismo a Singolo Nucleotide), è invece molto limitante nel sequenziamento di interi genomi.

### 2.2.2 Sequenziamento a colorante

Questo metodo, sfrutta l'amplificazione a ponte e interruzioni di sintesi reversibili, può essere suddiviso in tre fasi principali: amplificazione, sequenziamento ed analisi. Il filamento da determinare viene tagliato in segmenti più piccoli, che vengono poi posti su un particolare chip nel quale avviene l'amplificazione. Su questo chip sono presenti supporti solidi formati da oligo-nucleotidi, ai quali si agganciano i frammenti di DNA tramite indici (in genere di 6 nucleotidi) aggiunti precedentemente. A questo punto inizia l'amplificazione a ponte, un metodo che permette di creare centinaia di copie dei segmenti agganciati in poco tempo. In seguito alcuni metodi specifici applicano anche l'amplificazione per cloni, per un aumento dell'accuratezza e della qualità del sequenziamento, ma non è cruciale ai fini del processo stesso. Pulito il chip da tutto a parte che i filamenti "dritti" (cioè uguali agli originali), inizia la sintesi con basi marcate da molecole fluorescenti, che vengono aggiunte una alla volta. Il procedimento si ferma dopo ogni aggiunta in modo reversibile, viene scattata una foto e riprende dopo il lavaggio dell'ambiente dalle basi residue (come nel pirosequenziamento). Questo iter viene ripetuto per tutte le basi e per tutti i segmenti in cui era stato spezzato il campione. Le foto vengono poi analizzate da software per determinare la sequenza, assemblando tutte le rivelazioni ottenute, tramite confronti con campioni di riferimento.

Come nel caso precedente anche questo metodo è limitato dalla lunghezza del campione da leggere, inoltre la determinazione ex novo di una sequenza è più difficoltosa e risente grandemente della presenza di ripetizioni nella fase di assemblaggio durante l'analisi.

Un esempio concreto di applicazione per questo metodo è "Illumina" e in una sua versione leggermente diversa "SOLiD".

## Capitolo 3

# Sequenziamento tramite nanopori

Fra i metodi di sequenziamento di nuova generazione, ci concentreremo su quello tramite nanopori, che a sua volta può essere declinato in varie forme.

L'idea che sta alla base di questa tecnica, è che un singolo filamento di DNA, possa essere guidato, tramite elettroforesi, attraverso un nanoporo, ovvero un foro in una membrana isolante (o semiconduttrice in certi casi), dal diametro leggermente maggiore delle dimensioni del filamento ( 1.8-2 nm). Una soluzione viene divisa in due parti da questa membrana, in ognuna delle quali viene posto un elettrodo. In assenza del poli-nucleotide, il sistema crea una corrente di ioni attraverso il canale creato dal poro, generando una corrente di base  $I_0$ . A causa della carica negativa presente sui gruppi fosfati, che costituiscono l'impalcatura del filamento, questo viene attratto dal catodo. Infine il poli-nucleotide viene catturato ed entra nel poro. Durante un intervallo di tempo  $t_d$ , la durata della traslocazione, il filamento blocca parzialmente il flusso di ioni. Nonostante la carica negativa dei gruppi fosfati, il loro contributo in corrente è molto piccolo, poiché si muovono lentamente rispetto agli ioni liberi, ciò induce un calo di corrente fino ad un valore  $I_b$ , la corrente di ostruzione, che permette di rivelare la traslocazione del poli-nucleotide, più alcune informazioni su di esso, per esempio la sua lunghezza, o, in modo limitato, sulla sua composizione e dinamica.

La determinazione della sequenza avviene durante questo passaggio, secondo diversi approcci che ora citeremo ma che verranno poi analizzati in dettaglio: studio della corrente di ostruzione  $I_b$ , studio della corrente trasversa generata per effetto tunnel, impiego

dell'exonuclease e siti di rivelazione multipli per migliorare la precisione.

### 3.1 Traslocazione del DNA

Per analizzare il processo di traslocazione di un polinucleotide attraverso un poro, dobbiamo tenere conto di determinate proprietà, alcune generali riguardanti la dinamica dei polimeri (forze entropiche, moto Browniano, cariche, schermature); altre specifiche che dipendono dalla composizione atomica dei nucleotidi (ad esempio i potenziali di interazione con la superficie del poro).

#### Proprietà generali

I due processi alla base della dinamica dei polimeri carichi, sono la cattura e la traslocazione. La cattura dipende, dalla diffusione del polimero dal sito di deposito al poro e da effetti locali a ridosso del poro, per esempio il campo elettrico e l'interazione fra l'ingresso del poro ed il polimero. Il tasso di cattura inoltre, dipende dalla tensione applicata agli elettrodi, da quale molecola viene studiata, ed influenzerà la capacità di rivelare e sequenziare il campione. La traslocazione attraverso il poro viene azionata dalla differenza di tensione fra gli elettrodi, ma dipende da vari fattori: interazione polimero-poro, attrito viscoso ed effetti ionici. Tuttavia, due proprietà sono comuni alla traslocazione nei nanopori, ovvero: la carica effettiva e la schermatura dei polimeri nel poro e il problema correlato, cioè dove cade la tensione applicata.

In assenza di un polimero all'interno del poro, si nota che la maggior parte del calo di tensione fra le due sezioni avvenga nel poro, poiché questo ha una resistenza maggiore della soluzione circostante. Tuttavia, a seconda della sua forma (che dipenderà anche dalla natura del poro, biologica o sintetica), ci si aspetta che la presenza di un polimero al suo interno possa variare significativamente la caduta di tensione del sistema. Per fare un esempio, consideriamo il poro  $\alpha$ HL (che riprenderemo più avanti). Questo presenta, nella sua sezione finale, un canale più stretto (2 nm di diametro), ed è proprio in questo punto che, in presenza del polimero, il potenziale cala più drasticamente. A causa di ciò, ci si potrebbe aspettare che la tensione applicata agisca maggiormente proprio sui nucleotidi (o monomeri) presenti in questo segmento di poro. Una prova indiretta

di questa proprietà è data dal fatto che la dipendenza della velocità di traslocazione dal numero di nucleotidi cambia comportamento intorno ai 12 nucleotidi. Al di sopra di questo numero, la velocità del filamento è indipendente dalla sua lunghezza, ciò ci suggerisce che la forza motrice agisca solo su un numero finito di nucleotidi. Al di sotto dei 12 nucleotidi, la velocità cresce, in modo non lineare, con il calare del loro numero. Questo numero particolare, coincide con la lunghezza del segmento più stretto del filamento, dunque possiamo concludere che la forza che muove il filamento agisce su un numero finito di unità, che sono quelle contenute all'interno del poro.

Analizziamo ora l'interazione poro-polimero nel caso in cui il diametro di uno sia paragonabile con lo spessore dell'altro (casi in cui il diametro del poro è molto maggiore dello spessore del polimero sono poco interessanti ai fini del sequenziamento). In queste condizioni si ha una forte interazione, che aumenta l'attrito dinamico del filamento all'interno del poro. Il tempo di traslocazione  $t_d$  del polimero dipende linearmente da questo attrito, che, in questa configurazione, sovrasta gli altri fattori in gioco, come lo svolgimento del polimero (per filamenti di lunghezza molto maggiore rispetto allo spessore del poro).

Consideriamo ora il fatto che, ad essere in grado di passare attraverso il nanoporo, potrebbe non essere soltanto un polimero disteso linearmente. Se il diametro del poro  $D_p$  è sufficientemente largo da poter contenere più di un filamento, si potrebbero avere eventi corrispondenti a polimeri ripiegati. Nel nostro caso, analizziamo le forze agenti sulle coppie di nucleotidi del DNA, e di DNA ripiegato, che si trovano all'interno del poro. Se, come già detto, la tensione cala solamente all'interno del poro, la forza elettrica agente sul polimero in una regione data sarà:

$$F_{el}(m) \approx mF_{el}^0 \quad (3.1)$$

dove  $m$  è il numero di ripiegamenti (spire) nella regione considerata,  $F_{el}^0$  la forza agente su una singola coppia di nucleotidi,

$$F_{el}^0 = z_{eff}E \quad (3.2)$$

con  $z_{eff} \approx -0.5e$  carica effettiva sulla coppia di nucleotidi ed  $E$  campo elettrico. Se l'interazione poro-polimero è la principale responsabile della forza di attrito, allora

otterremo una relazione simile alla prima,

$$F_{att}(m) \approx mF_{att}^0 \quad (3.3)$$

dove  $F_{att}^0 \approx \eta v$ , con  $\eta$  coefficiente proporzionale alla superficie di contatto fra poro e polimero, da cui  $\eta \rightarrow \eta v$  per il polimero ripiegato, e  $v$  è la velocità del polimero. Assumendo  $v$  costante, avremo che

$$F_{el} = -F_{att} \quad (3.4)$$

Per cui i fattori  $m$  si semplificano e  $v$  rimane invariata per polimeri ripiegati o distesi ( $v = v_0$ ). Consideriamo ora un polimero con una regione di lunghezza  $L_m$ , contenente  $m$  ripiegamenti, con  $L_m \gg L_p$ . Se consideriamo l'occupazione del volume del poro come causa dell'ostruzione ionica, la corrente d'ostruzione del polimero disteso e ripiegato, saranno legate dalla seguente legge

$$I_0 - I_b^m = m(I_0 - I_b^0) \quad (3.5)$$

dove  $I_0$  è la corrente con il poro aperto,  $I_b^0$  è la corrente di ostruzione per un polimero disteso e  $I_b^m$  per uno con  $m$  ripiegamenti. Dalla (3.5) si deduce che la carica totale bloccata per la traslocazione di una regione  $L_m$  di polimero ripiegato, è la stessa di una regione di lunghezza  $mL_m$  di un polimero disteso, che contiene lo stesso numero di nucleotidi,

$$(I_0 - I_b^m) \frac{L_m}{v} = m(I_0 - I_b^0) \frac{mL_m}{v_0} \quad (3.6)$$

chiamata regola (ECD), difetto di carica d'evento costante,

$$(ECD) \equiv \Delta I_b t_b \quad (3.7)$$

dove  $\Delta I_b = I_0 - I_b$  e  $t_b$  è il tempo di traslocazione. Dai dati sperimentali emerge però una discrepanza più o meno marcata dalle curve d'ECD costanti. Questo è dovuto principalmente alle approssimazioni fatte sulle forze d'attrito nel caso di pori di diametro vicino allo spessore del polimero, oltre che ad altri effetti casuali, come ds-DNA che può temporaneamente legarsi alle pareti del poro, o il distendersi del ds-DNA all'esterno.

## Proprietà specifiche

Una proprietà specifica per noi interessante, notata durante esperimenti con pori  $\alpha$ HL, è che la velocità varia anche a seconda del verso di traslocazione, cioè il filamento può entrare nel poro orientato  $3' \rightarrow 5'$  o  $5' \rightarrow 3'$ . La diversa orientazione influisce sulla velocità poiché le basi azotate non sono perpendicolari all'impalcatura, dunque cambia l'angolo fra loro e la superficie del poro, da cui dipende l'interazione polinucleotide-superficie.

## Tipi di nanopori e loro caratteristiche

I nanopori che costituiscono l'apparato di sequenziamento, possono essere suddivisi in due macro-categorie in base alla loro natura: nanopori sintetici e nanopori biologici. Dei primi parleremo poco, in quanto sono poco usati solitamente, mentre della seconda categoria parleremo più nel dettaglio, concentrandoci in particolare su due casi specifici:  $\alpha$ HL e MspA.

Le caratteristiche del poro e la possibilità di controllarle sono cruciali per i meccanismi di sequenziamento. Per esempio, metodi che prevedono l'uso di sonde nanometriche incorporate nel poro, richiederanno che esso abbia dimensioni ottimali per massimizzare la differenza di segnale fra le diverse basi, in modo da sovrastare le varie, inevitabili, fonti di rumore. Questa situazione si traduce in un diametro del poro molto vicino allo spessore di un singolo filamento di DNA. Tuttavia la massimizzazione del rapporto segnale-rumore, deve essere bilanciata da altri effetti, come l'interazione DNA-superficie, che è minima per pori di grandi dimensioni, o anche la cattura e la traslocazione del filamento, che non avviene per pori sottodimensionati.

### 3.2 Nanopori sintetici

I nanopori sintetici offrono alcuni vantaggi rispetto ai pori biologici. Si possono gestire più liberamente determinate condizioni di lavoro del poro, per ottenere il massimo rendimento del sequenziamento, come concentrazione salina, temperatura, tensione, viscosità; possiamo creare pori con le dimensioni ottimali per l'esperimento, con una buona pre-



cisione, incorporarvi ulteriori strumenti di misura, ad esempio elettrodi che misurino la corrente trasversa durante il passaggio del poli-nucleotide. Riportiamo ora due tecniche di costruzione per questo tipo di pori.

La prima prevede la creazione di un foro di grande diametro ( 60 nm) in una membrana allo stato solido di  $Si_3N_4$ . Questo foro può essere ottenuto bombardando la membrana con un raggio focalizzato di ioni. A questo punto, il foro è esposto ad un raggio di ioni  $Ar^+$  a bassa energia, che non rimuove atomi dalla membrana allargando ulteriormente il poro come si potrebbe pensare, ma innesca un processo di diffusione che porta il foro a restringersi. La corrente di  $Ar^+$  in uscita dal poro, dipende linearmente dalla superficie dello stesso, dunque misurandola come funzione del tempo, si può controllare il restringimento fino alla scala nanometrica. Un difetto di questa tecnica è che una volta ristretto il poro, non si conosce la concentrazione di silicio ed azoto sulla sua superficie interna. Ciò può portare alla presenza di zone elettricamente cariche che influenzerebbero la traslocazione dei poli-nucleotidi.

La seconda è concettualmente simile alla prima, cioè è basata sull'idea di restringere il diametro del poro sottoponendolo ad irradiazione. In questo caso, la membrana è costituita da  $SiO_2$ , nella quale il foro più largo di partenza può essere creato in vario modo. Una volta presente, questo viene ristretto utilizzando un raggio elettronico ad alta energia, che può essere usato anche come microscopio elettronico a trasmissione (TEM). Sfruttando la capacità di restituire immagine del TEM, si può osservare il poro restringersi. Siccome la velocità di restringimento è bassa ( 0.3 nm al minuto in queste condizioni), si possono usare le immagini per monitorare il processo ed arrestarlo una volta raggiunto il diametro voluto. Il limite sulla precisione raggiungibile è dato dalla risoluzione del microscopio ( 0.2 nm). Tuttavia a causa dell'irregolarità della superficie, il diametro può essere controllato solamente fino a 1 nm. La silice, per i nostri scopi, ha il difetto di avere una densità di carica superficiale in soluzione acquosa.

L'osservazione che raggi ad alta energia, possono innescare un processo di restringimento nei pori, rimane un fenomeno interessante in sè, oltre alle implicazioni che ha nella creazione di pori e strutture alla scala nanometrica.

### 3.3 Nanopori Alpha emolisina

#### 3.3.1 Caratteristiche del nanoporo

L' $\alpha$ -emolisina ( $\alpha$ HL) è una proteina costituita da sette sotto unità identiche fra loro. Ha un canale centrale di diametro variabile (14-26 Å), rispetto al quale ha simmetria assiale. E' suddivisibile in due sezioni, separate da un restringimento (14 Å), entrambe lunghe circa 50 Å, la prima più larga, che costituisce il vestibolo dal quale entra il filamento; la seconda, più stretta, costituisce l'ambiente dove avviene la discriminazione dei nucleotidi.

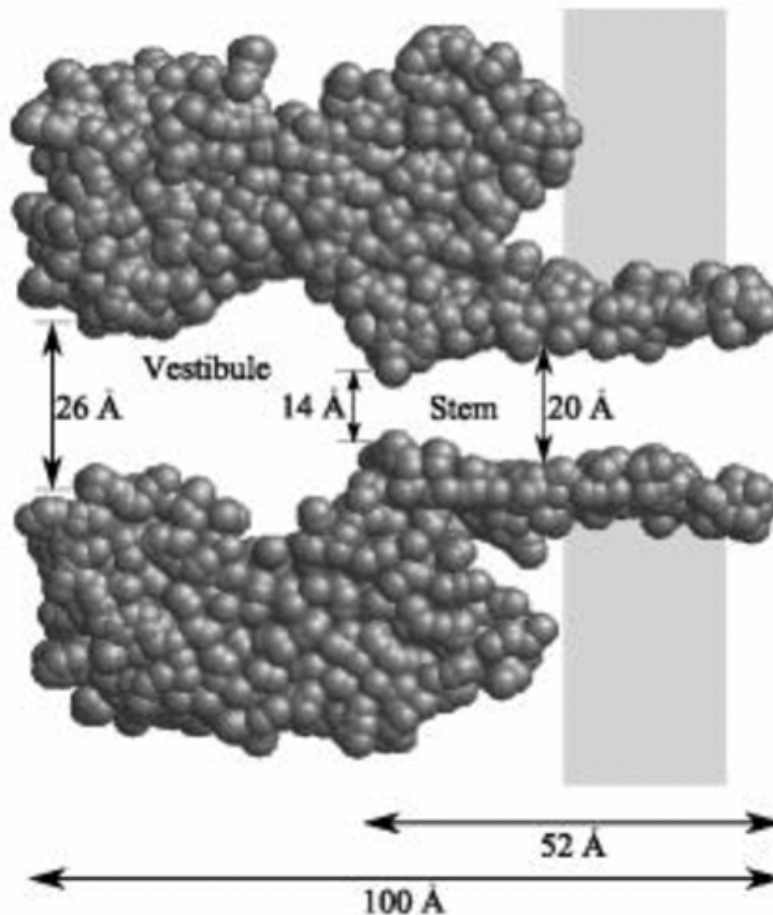


Figura 3.1: Schematizzazione del nanoporo Alpha emolisina inserito in una membrana

In questo poro biologico, la restrizione più piccola, attraverso cui deve traslocare il

polinucleotide, ha un diametro di 1.4 nm, dunque il ds-DNA, con un diametro di 2 nm, non è in grado di passarvi attraverso, mentre può il ss-DNA. Inoltre il foro è sia stretto sia sufficientemente lungo così che il filamento debba essere localmente disteso; ciò crea una barriera entropica per il trasporto a causa dello svolgimento del polimero.

Il canale finale, di diametro quasi costante, può essere modificato a seconda del metodo di sequenziamento utilizzato. Ad esempio, possono essere implementati elettrodi, per discriminare i nucleotidi utilizzando una corrente trasversa dovuta all'effetto tunnel.

### 3.3.2 Corrente ionica di ostruzione

La corrente di ostruzione è in parte dovuta all'"esclusione volumica". Il polinucleotide occupa una parte del volume del poro, impedendo agli ioni, quindi ai portatori di carica di passare, dunque riduce la corrente ionica misurata. Quanto volume viene occupato dal polinucleotide dipende dalla sua struttura e composizione, ad esempio polinucleotidi con una struttura secondaria elicoidale, occupano una porzione di volume maggiore, se riescono a mantenerla intatta. Dovremo poi tenere conto di quanto altri eventuali fattori possano influire sulla corrente, come l'interazione DNA-poro e l'elettrostatica del sistema.

Per farci un'idea del contributo dell'esclusione volumica allora corrente di ostruzione, consideriamo l'equazione stazionaria di Nernst-Planck per il flusso di carica  $J_z$  lungo la direzione  $z$

$$J_z = e\mu n E_z - eD \frac{\partial n}{\partial z} \quad (3.8)$$

dove  $e$  è la carica elettrica (assumendo come portatori di carica cationi monovalenti),  $\mu$  la mobilità ionica,  $n$  la densità dei portatori di carica,  $D$  il coefficiente di diffusione ed  $E_z$  il campo elettrico dovuto alla tensione applicata. L'assunzione generale alla base dell'equazione è una distribuzione continua di ioni contenuta in un dielettrico continuo. La forma particolare della (3.8), assume che non ci sia una barriera di energia libera al passaggio degli ioni nel poro. In assenza di un qualche tipo di termine di diffusione, la corrente a poro aperto diventa

$$I_0 = e\mu n E_z \quad (3.9)$$

dove  $E_z \approx V/L_p$  (la tensione cala quasi totalmente sulla lunghezza del poro). La quantità  $e\mu n = \sigma$  è la conduttività del poro.

Se l'esclusione volumica fosse l'unico fattore nel calo di corrente, potremmo scrivere

$$n_b = \frac{V_p - V_N}{V_p} n = F n \quad (3.10)$$

dove  $n_b$  è la densità dei portatori di carica durante l'ostruzione,  $V_p$  è il volume del segmento di poro occupato da un nucleotide,  $V_N$  è il volume di un nucleotide. Dunque, per la riduzione relativa della corrente da  $I_0$  a  $I_b$  otteniamo

$$\frac{I_0 - I_b}{I_0} = 1 - F \quad (3.11)$$

Più semplicemente, se assumiamo che la densità di portatori sia la stessa, ma che semplicemente una parte dell'area del poro  $A_0$  sia stata ostruita, allora  $(I_0 - I_n)/I_0 = A_b/A_0$  con  $A_b$  la porzione di area ostruita. Questo richiede regolarità nell'orientazione delle unità, ma non sempre questo è totalmente corretto.

I dati sperimentali non sono però d'accordo con queste premesse teoriche, in particolare ci suggeriscono che il contributo al calo di corrente non sia solamente dovuto all'esclusione volumica, ma anche ad altri fattori dipendente ed indipendenti da questo. Ad esempio l'esclusione di volume crea un'ulteriore barriera elettrostatica a causa dell'ulteriore confinamento degli ioni (mentre precedentemente avevamo assunto la densità dei portatori costante), "disattivando" sostanzialmente dei portatori di carica. Inoltre non abbiamo nemmeno considerato il fatto che il polinucleotide stesso trasporta una quantità (seppur piccola) di carica all'interno del poro. Abbiamo poi effetti che trascuravamo come l'interazione poro-polinucleotide, che a questo punto giocherà un ruolo rilevante.

### Proprietà elettrostatiche

Un altro problema legato alla corrente ionica, è come le proprietà elettrostatiche del poro influiscono sul trasporto. Una prima, parziale, risposta può essere data considerando l'equazione di Nernst-Planck completa, con un modello semplificato delle proprietà elettrostatiche del poro. Introducendo dunque delle appropriate costanti dielettriche e mobilità ioniche dipendenti dalla posizione, un potenziale opportuno per l'energia libera. Quest'ultimo sarà tanto più rilevante quanto più piccolo sarà il diametro del poro,

essendo più difficile il trasporto delle cariche. Dunque nel caso in cui la maggior parte degli ioni non abbia energia sufficiente per attraversare il poro (cioè con grandi variazioni di energia libera), consideriamo un processo attivato in cui il numero degli ioni effettivamente disponibili per il trasporto è dato da

$$n = n_0 e^{-\Delta F_b/kT} \quad (3.12)$$

dove  $n_0$  è la densità alla sorgente,  $\Delta F_b$  rappresenta la barriera di energia libera, dovuta sia alla barriera di energia elettrostatica che alla barriera entropica. Per pori di larghe dimensioni sono entrambe trascurabili.

## Risultati

Le diverse basi restituiscono effettivamente diversi segnali ed impiegano tempi diversi a traslocare attraverso il poro, ma ciò è apprezzabile solamente in omopolimeri (cioè polimeri composti da un unico tipo di monomero). Non è possibile sequenziare polinucleotidi costituiti da basi diverse. È anche possibile distinguere filamenti di omo-RNA da filamenti di omo-DNA, grazie alle diverse strutture secondarie che influiscono sia su  $I_b$  che su  $t_d$ .

Dunque la misura  $I_b$  permette di rivelare eventi di traslocazione e stabilirne il verso,  $3' \rightarrow 5'$  o  $5' \rightarrow 3'$ , discriminare le basi per polinucleotidi omogenei o con lunghe sequenze omogenee. Per poter sequenziare filamenti eterogenei, lo spessore del poro dovrebbe coincidere con l'estensione del un monomero, ma questo creerebbe altre complicazioni. Si potrebbero usare nanopori appositamente ingegnerizzati con molecole che abbiano diverse interazioni con ogni base, modificandone  $I_b$  caratteristica, o con enzimi come l'exonuclease per analizzare un nucleotide per volta. In entrambi i casi perderemmo molto in velocità di sequenziamento.

### 3.3.3 Corrente trasversa

Alla scala nanometrica, gli elettroni possono dare effetto tunnel, cioè essi hanno una probabilità di penetrare regioni proibite dall'elettrodinamica classica, generando una corrente elettronica. Questo fenomeno avviene in qualunque mezzo (vuoto, fluidi, mo-

lecole etc.), basta che l'intervallo energetico sia sufficientemente piccolo. La probabilità che si verifichi tuttavia, dipende dal mezzo, in particolare le molecole forniscono livelli energetici, che aumentano la corrente elettronica a seconda di quanto sono distanti dal livello di Fermi degli elettrodi metallici e in che modo si accoppiano ad essi.

Questo fenomeno, può essere sfruttato per distinguere le diverse basi, che hanno strutture molecolari distinte, quindi delle proprie caratteristiche elettroniche. Effettivamente le basi hanno una diversa corrente di trasporto caratteristica. Questa però, è influenzata in grande misura dall'accoppiamento base-elettrodo, che a sua volta risente della posizione del nucleotide e dei suoi movimenti, determinati dalla traslocazione del filamento attraverso il nanoporo, dall'interazione nucleotide-fluido nello spostamento e anche da movimenti intrinseci del polinucleotide. La soluzione a questo problema può essere, oltre a correzioni dei dati raccolti con opportuno algoritmi, l'applicazione di un campo trasversale agli elettrodi. Con un condensatore ad esempio, si può creare un campo elettrico perpendicolare all'asse degli elettrodi che attiri l'impalcatura del PN distendendone le basi.

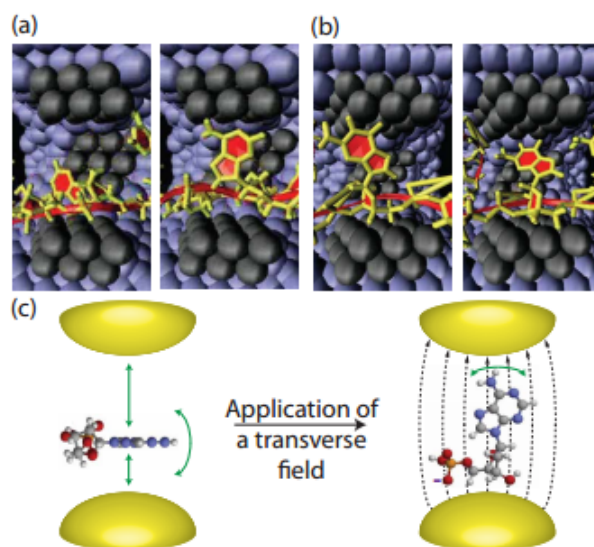


Figura 3.2: Simulazione di dinamica molecolare della traslocazione di un polinucleotide attraverso un nanoporo. Senza un campo trasversale (a) si nota che le fluttuazioni nell'orientazione delle basi sono grandi, con un campo trasversale (b) diminuiscono di molto. (c) schematizzazione di un nucleotide fra due elettrodi in assenza e in presenza del campo trasversale

Con questo accorgimento, le distribuzioni di corrente ottenute per ogni base, sono approssimativamente log-normali e presentano solo sovrapposizioni parziali, il che le rende distinguibili con un basso numero di prese dati indipendenti. Il loro aspetto fondamentale (distribuzione gaussiana in scala logaritmica delle fluttuazioni di corrente) è dovuto alla dipendenza esponenziale dell'accoppiamento elettronico dalla distanza elettrodo-nucleotide. Allo stesso modo possiamo pensare al trasporto attraverso la giunzione, come dovuto ad un unico stato elettronico sul DNA. Con queste considerazioni possiamo approssimare la corrente trasversa nel seguente modo

$$I(\eta) \approx \frac{2e^2V}{h} \frac{\gamma^2}{E_0^2 - \eta^2} \quad (3.13)$$

dove è stata assunta una risposta lineare, che dà una proporzionalità diretta della corrente dalla tensione applicata  $V$ .  $e$  è la carica dell'elettrone,  $h$  la costante di Planck. Questa è l'equazione approssimata per la corrente per un unico livello elettronico di energia  $E_0$ , accoppiato allo stesso modo ad entrambi gli elettrodi con una forza  $\gamma$  ed in presenza di rumore di intensità  $\eta$ .

Quest'espressione tuttavia si riferisce ad un singolo istante di tempo, mentre, come già menzionato, sappiamo che l'interazione DNA-elettrodi varia. Dunque la costante d'accoppiamento corretta sarà espressa dalla sua distribuzione di probabilità

$$p(\ln \gamma/\gamma_m) = \frac{1}{\sqrt{2\pi\sigma_\gamma^2}} \exp - \frac{(\ln \gamma/\gamma_m)^2}{2\sigma_\gamma^2} \quad (3.14)$$

con deviazione standard  $\sigma_\gamma$  e massimo valore di aspettazione  $\gamma_m$ . È importante ricordare, che questi risultati sono validi solamente nel caso in cui il moto del polinucleotide sia controllato tramite un campo trasversale, che attira l'impalcatura verso uno degli elettrodi. Le distribuzioni per le correnti caratteristiche di ogni base sono riportate in figura (3.3).

## Risultati

La discriminazione, essendo le correnti date da delle distribuzioni, non è possibile con una singola misura, ma solo con più misure, o acquisendo la media della corrente sul tempo di misura di un singolo nucleotide presente fra gli elettrodi.

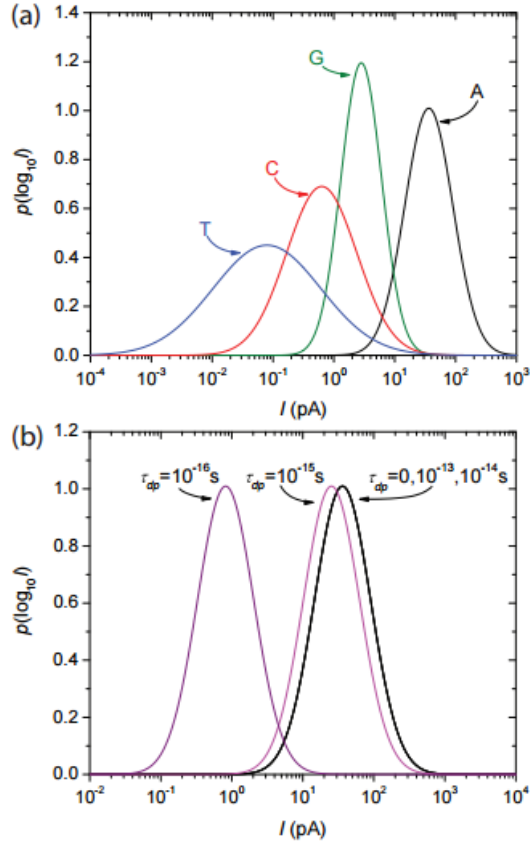


Figura 3.3: (a) Distribuzioni delle correnti elettroniche caratteristiche di ogni base posta fra due elettrodi d'oro con rumore  $\eta=0$ . (b) Variazione delle distribuzioni dell'adenina in presenza di rumore dovuto a diversi sfasamenti degli elettrodi  $\eta = \hbar/\tau_{dp}$ . Le distribuzioni sono state calcolate con simulazioni di dinamica molecolare, con tensione 1 V e distanza fra gli elettrodi 1.4 nm

Ulteriori risultati sperimentali indicano inoltre, che questo tipo di misura non viene influenzato sensibilmente dall'ambiente ionico e acqueo. Anche il rumore bianco dovuto allo sfasamento dei due elettrodi, di cui si è già tenuto conto nella (3.13), ha effetti trascurabili per intensità moderate. Solamente intensità maggiori dell'intervallo energetico fra il livello elettronico del nucleotide e il livello di fermi degli elettrodi creeranno disturbi apprezzabili. Quindi in generale il sistema di misura è schermato da questo tipo di rumore.



### 3.3.4 Siti di rivelazione

Ci siamo già posti il problema di capire da quanti e quali nucleotidi viene influenzata la corrente ionica di ostruzione  $I_b$ . Cerchiamo ora di studiare il problema inverso, cioè a quali nucleotidi è sensibile il nanoporo  $\alpha$ HL in base allo studio di  $I_b$ . Per poter fare questo, è necessario rallentare o immobilizzare il DNA mentre si trova nel poro, altrimenti il tempo di traslocazione  $t_d$  dei singoli nucleotidi sarebbe troppo breve per poter fare misure significative. A tal scopo possiamo impiegare l'enzima DNA-polimerasi che rallenta il filamento e lo fa avanzare a scatti di un nucleotide, DNA con strutture secondarie note (forcine, avvolgimenti ad elica) oppure, per immobilizzarlo, una proteina chiamata NeutrAvidina (una streptavidina). Gli ultimi due metodi, danno risultati più netti, per il tempo d'osservazione prolungato.

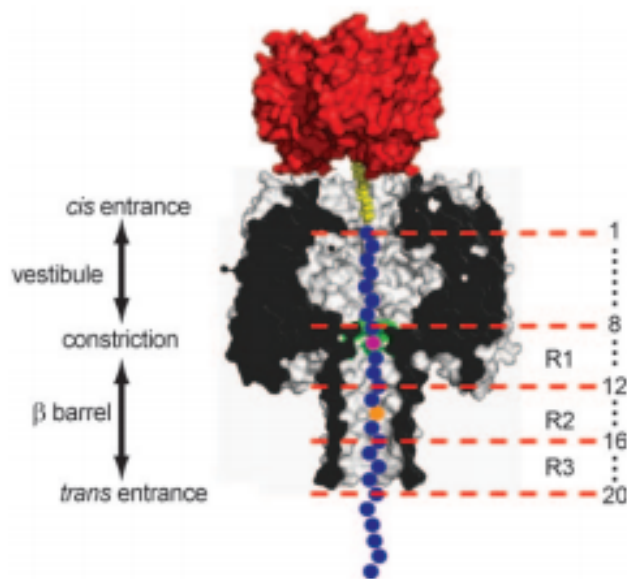


Figura 3.4: Schematizzazione del nanoporo Alpha-emolisina con all'interno un polinucleotide immobilizzato da NeutrAvidina. Sono riportati sulla destra i siti di rivelazione e le rispettive posizioni dei nucleotidi.

Adoperiamo una mutazione dell' $\alpha$ HL, dove sono stati sostituiti alcuni amminoacidi per diminuire l'interazione polinucleotide-poro in zone indesiderate. Introduciamo nel poro filamenti omo-DNA, con una base sostituita, per determinare la capacità di di-

stinguere le quattro basi e acquisirne le correnti residue caratteristiche. Si ottengono differenze fra i picchi delle distribuzioni molto piccole, dell'ordine del pA. Allo stesso modo, introducendo filamenti con un'unica base diversa, agganciati a NeutrAvidina, in modo che rimangano immobilizzati nel poro, possiamo determinare quante zone sensibili ci sono, e di quali nucleotidi risentono. I siti sono tre, R1, R2, R3, ognuno di essi risente della presenza di quattro nucleotidi, partendo a contare dalla streptavidina, R1 è sensibile alle posizioni 8-12, R2 da 12 a 16 e R3 da 17 a 20. Queste procedure sono spiegate in dettaglio nel paragrafo 3.4.2 per il nanoporo MspA e sono del tutto analoghe.

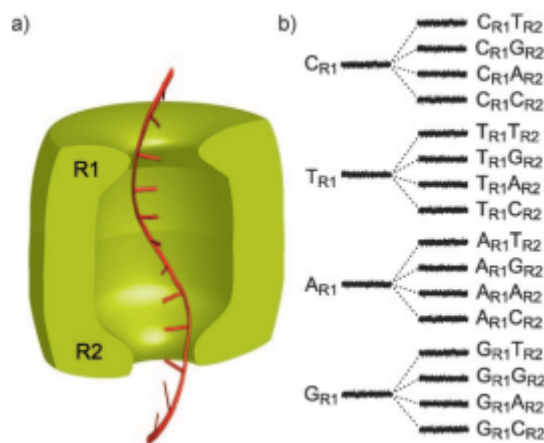


Figura 3.5: (a) Schematizzazione di un poro con due siti di rivelazione R1 e R2. (b) I possibili segnali di lettura dovuti alle diverse combinazioni di lettura di R1 e R2.

Sapendo ora quali sono i siti di rivelazione del poro, vogliamo capire come sfruttare al meglio le informazioni che possono fornirci. Pensiamo per ora di avere solamente due siti, le quattro basi produrranno quattro livelli di corrente distinti nel primo sito, nel secondo, potrà esserci una qualunque delle basi, quindi ogni livello di corrente precedente, sarà suddiviso in altri quattro, per un totale di sedici livelli. Quello che andremo a rivelare sarà dunque l'effetto di una di queste sedici combinazioni, dovremo quindi essere in grado di distinguere non più quattro segnali, uno per base, bensì sedici, uno per combinazione. I dati sperimentali finora, sono stati in grado di distinguere solamente undici segnali. Questo però non è un problema, in quanto ogni base verrà rivelata due volte, mentre trasloca attraverso R1 e poi attraverso R2, avremo quindi informazioni ridondanti su tutti i nucleotidi, migliorando di fatto la qualità della lettura. Considerando anche la terza

testa, si potrebbe pensare che avere una lettura in più migliori ancora il sequenziamento, ma così non è (al momento), perché le correnti da distinguere non sarebbero più sedici bensì sessantaquattro, un numero troppo alto per poter essere sufficientemente separato, dato il rumore del sistema. I siti di rivelazione di questo poro inoltre, non sono tanto acuti da poter distinguere un nucleotide per volta, come abbiamo ipotizzato, ma risentono della presenza di quattro nucleotidi per volta, quindi la misura diventa ancora meno precisa.

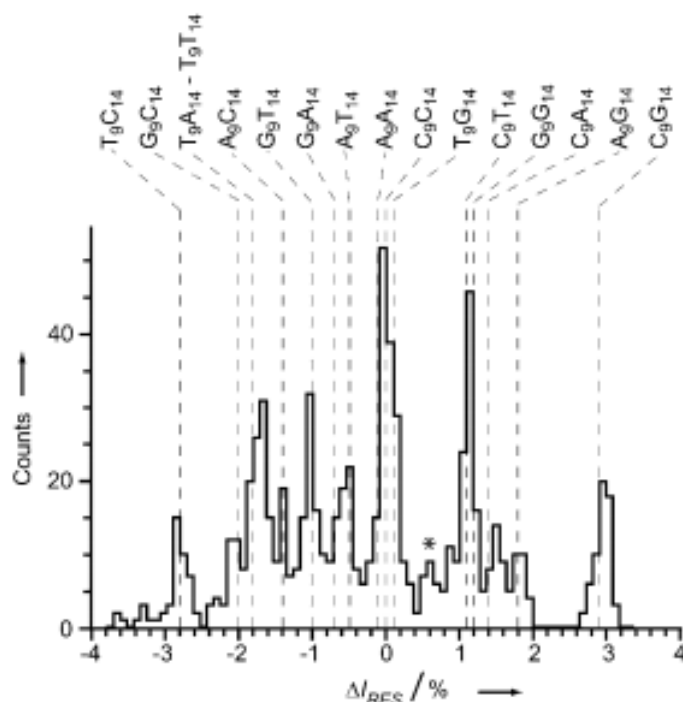


Figura 3.6: Istogramma delle correnti residue relative dovute alle diverse combinazioni nucleotidiche fra due siti di rivelazione

### 3.3.5 Risultati del sequenziamento tramite Alpha emolisina

I risultati mostrati dal sequenziamento del DNA tramite  $\alpha$ HL sono molto buoni e promettenti.

La corrente di ostruzione ionica non è ancora in grado di fornirci uno strumento sufficiente alla discriminazione, ma affinando le tecniche, in particolare con mutazioni della proteina che permettano di avere siti di rivelazione con una risoluzione sempre

maggiore, fino al singolo nucleotide, anche solo questa misura potrà essere determinata la sequenza.

Altri approcci, come la misura della corrente trasversa (electron tunneling), mostrano un grande potenziale, il loro ulteriore sviluppo porterebbe ad una velocizzazione ed ottimizzazione della tecnica. Questo criterio potrebbe essere applicato anche ad altri pori in cui sia possibile inserire degli elettrodi sufficientemente piccoli.

Tramite l'impiego di enzimi come la DNA-polimerasi o l'exonuclease, è già possibile sequenziare lunghi filamenti in tempi relativamente brevi. Il progresso della tecnica permetterebbe di eliminare questi enzimi, in favore di un controllo della traslocazione tramite campi elettrici adeguati ai tempi di misura all'interno del poro, riducendo i quali si velocizzerebbe enormemente il sequenziamento.

## 3.4 Nanopori *Mycobacterium smegmatis* porin A

### 3.4.1 Caratteristiche del nanoporo

Il *Mycobacterium smegmatis* porin A (MspA) è una proteina formata da otto sotto unità identiche fra loro. Ha un canale centrale di diametro variabile, rispetto al quale ha simmetria assiale. Il diametro cala (da quello che consideriamo l'ingresso del canale) quasi uniformemente fino a raggiungere 12 Å nella parte terminale, lunga 6 Å. Tuttavia in questo sito sono presenti cariche negative che respingono l'impalcatura ed impediscono al polinucleotide di passare. Viene dunque impiegata M1-MspA, una mutazione del MspA naturale, senza cariche superficiali. Queste dimensioni ne fanno una candidata ideale come nanoporo per il sequenziamento del DNA.

### 3.4.2 Corrente ionica e siti di rivelazione

Ricordando le considerazioni fatte sulla corrente ionica di ostruzione per l' $\alpha$ HL, notiamo come l'estensione del restringimento finale del MspA (6 Å) sia ottimale per l'acquisizione di  $I_b$  nucleotide per nucleotide. Con queste dimensioni infatti, se effettivamente la zona sensibile del poro è solamente quella più stretta, il segnale che otteniamo è dovuto per la maggior parte unicamente al nucleotide ivi contenuto in quel momento.

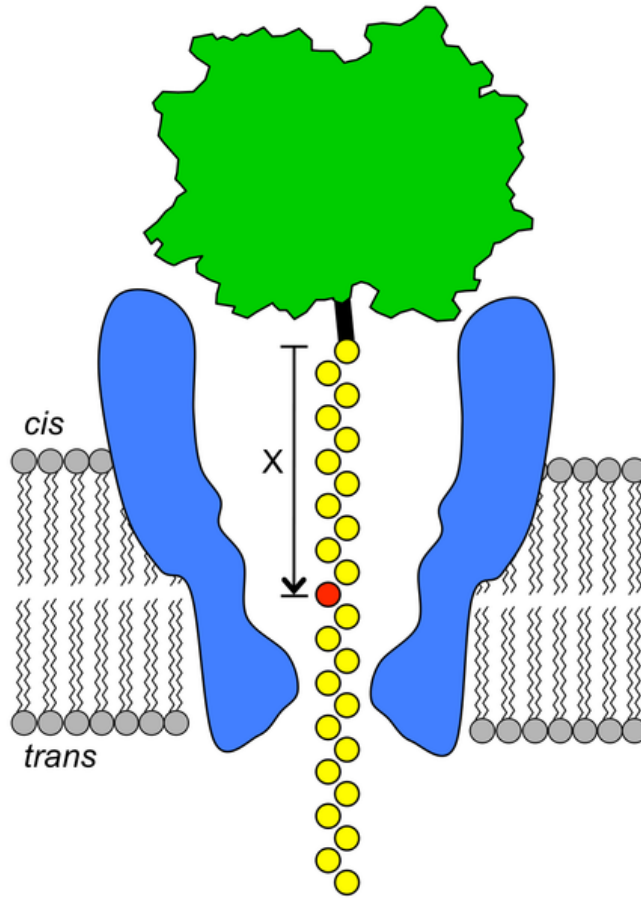


Figura 3.7: Disegno schematico in sezione, di un nanoporo MspA inserito in una membrana lipidica, con all'interno un polimero immobilizzato agganciato ad una NeutrAvidina

Si presentano ora due problemi: determinare la zona di sensibilità del poro e la velocità di passaggio (e quindi  $t_d$ ) in funzione della tensione applicata. Partendo dal secondo, per DNA libero la traslocazione è molto veloce  $1-7 \mu\text{s}/\text{nucleotide}$ , troppo per poter distinguere il segnale dal rumore di fondo. Abbiamo diverse possibili soluzioni, diminuire la tensione applicata, ma questo diminuirebbe anche il segnale acquisito, lasciando invariato il SNR; usare DNA modificato, ad esempio con segmenti a doppio filamento, che dovrebbero essere rimossi per poter traslocare; infine potrebbero essere impiegati motori molecolari, che facciano avanzare il filamento una base alla volta, ad esempio specifiche DNA-polimerasi o streptavidine, in particolare NeutrAvidina.

Date queste precisazioni iniziali passiamo allo studio di  $I_b$ . Per prima cosa si acqui-

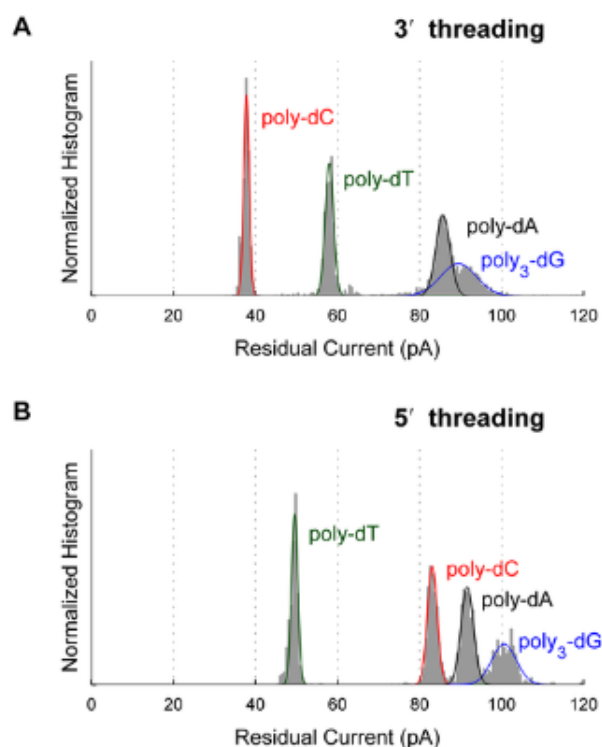


Figura 3.8: Istogrammi e fit gaussiani normalizzati delle correnti residue caratteristiche dei diversi polinucleotidi omogenei immobilizzati nel nanoporo MspA, agganciati a NeutrAvidina in posizione 5' (A) e 3'(B)

siscono le correnti caratteristiche di filamenti di omo-DNA, operando la traslocazione in entrambe le direzioni  $3' \rightarrow 5'$  e  $5' \rightarrow 3'$ . I risultati ottenuti hanno mostrato che la presenza della proteina, che funge da motore molecolare, non incide sensibilmente sulla corrente ionica misurata. In figura (3.8) sono mostrati gli istogrammi delle correnti residue medie, con relative curve gaussiane di fit. Notiamo che i segnali sono ben distinti, con una separazione minima di 8 pA, per il verso  $3' \rightarrow 5'$  c'è una sovrapposizione del 27% fra A e G, che sparisce quasi totalmente nel verso contrario. Ciò è dovuto probabilmente ad esigenze sperimentali, che impediscono l'utilizzo di un filamento poli-dG, il segnale associato alla guanina è in realtà dovuto a un filamento poli-dA in cui viene inserito un segmento poli-dG. È evidente anche la dipendenza di  $I_b$  dal verso di traslocazione, la corrente residua dei filamenti poli-dC varia di 45 pA.

Una richiesta dei metodi di sequenziamento, è la capacità di distinguere le basi na-

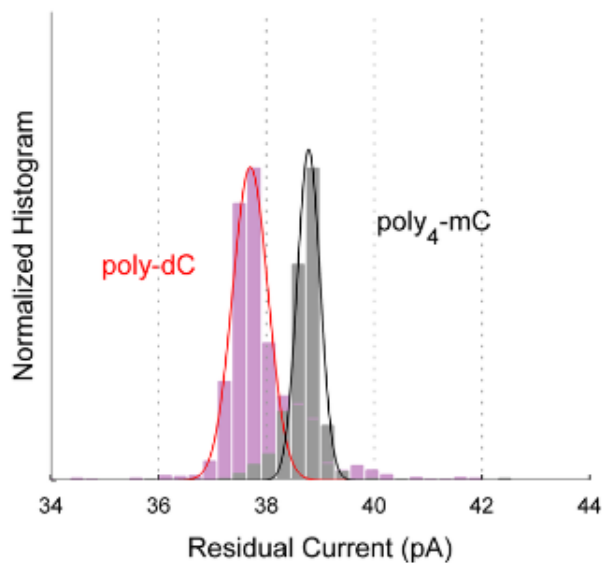


Figura 3.9: Istogrammi e fit gaussiani normalizzati delle correnti residue caratteristiche della policitosina e della poli-metil-citosina

turali da basi metilate, cioè base con una particolare variazione che svolge importanti funzioni biologiche. Nello specifico interessa essere in grado di distinguere poli-dC e poli-mC. Il grafico in figura (3.9), ottenuto con le stesse procedure delle misure precedenti, mostra una separazione dei valori medi di 1.1 pA, con una sovrapposizione del 2%.

Vogliamo ora determinare quale sia la regione di sensibilità del poro, se davvero come abbiamo ipotizzato è solamente il restringimento finale, o se anche in altri punti la corrente cala in presenza di nucleotidi. Per verificarlo, adoperiamo un filamento poli-dA nel quale inseriamo una singola dC progressivamente nelle posizioni da 11 a 18 (contando dalla NeutrAvidina) e manteniamo il filamento immobile nel poro. Queste posizioni sono dettate dalle dimensioni del poro e della NeutrAvidina, oltre il ventesimo nucleotide infatti il filamento si trova già nella camera trans-membrana. I risultati sono riportati nel grafico in figura (3.10a). Assumendo una distribuzione gaussiana dei dati ottenuti, la regione sensibile si trova in  $X=14.5$  con larghezza a mezza altezza di 1.5 nucleotidi. Ripetiamo il processo usando il filamento con le basi invertite, cioè poli-dC con una sola dA, viene usato anche un poli-dT con una sola dA, poiché nel primo caso non abbiamo variazioni utili. Il grafico in figura (3.10b) mostra un fit gaussiano, con valor medio 14.1 e larghezza a mezza altezza 1.5 nucleotidi. Ora sappiamo che la regione sensibile

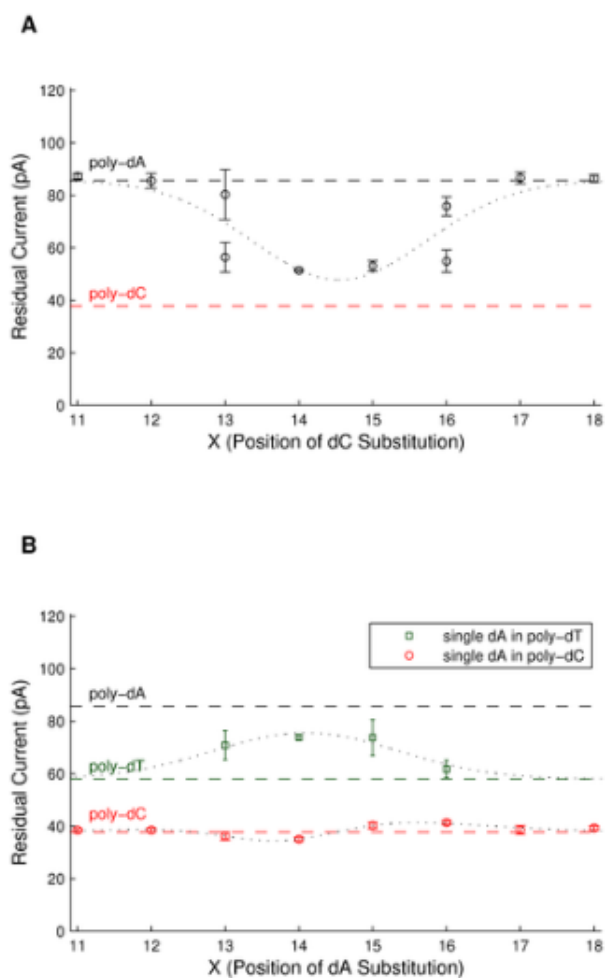


Figura 3.10: Valori di corrente residua in omo-polimeri con singola sostituzione nella posizione indicata dall'ascissa

coincide con il restringimento del poro, che vi si trovano i nucleotidi in posizione 14-15 (utilizzando NeutrAvidina per rallentare la traslocazione) e che il MspA è sensibile alla mutazione di almeno un singolo nucleotide in un filamento omogeneo.

Studiamo ora la capacità di rivelare una sostituzione in un filamento etero-nucleotidico. Data una sequenza nucleotidica, sostituiamo alternatamente dA e dC progressivamente dalla posizione 13 alla 16. Dai grafici (3.11) si nota che i segnali sono ben distinguibili quando la sostituzione si trova nelle posizioni 14 e 15, con una separazione di almeno 25 pA.



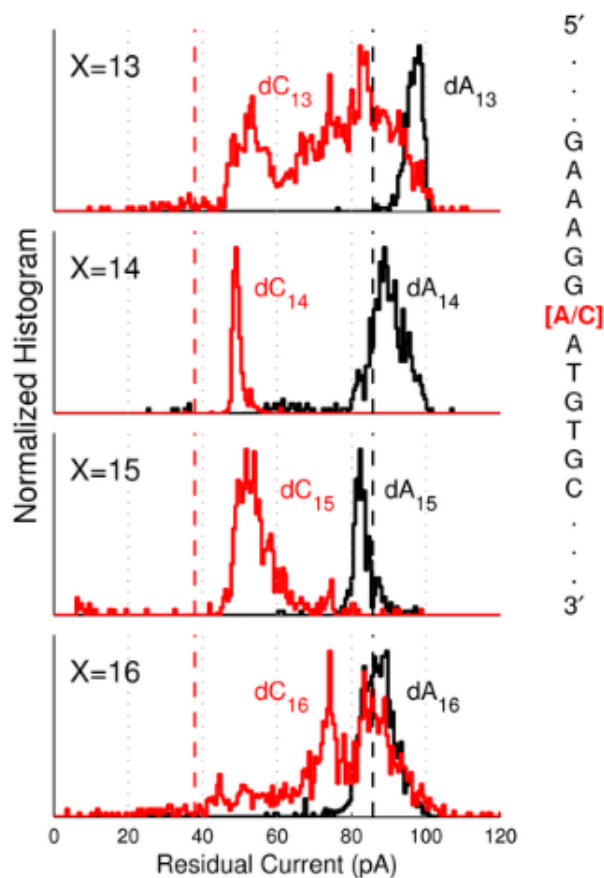


Figura 3.11: Istogrammi dei conteggi di corrente residua per sostituzioni singole di A e C in eteropolimeri noti, in quattro diverse posizioni X

### 3.4.3 Risultati del sequenziamento tramite *Mycobacterium smegmatis* porin A

Le differenze fra le correnti di ostruzione specifiche di ogni nucleotide, sono maggiori di circa un ordine di grandezza rispetto alle controparti ottenute con nanopori  $\alpha$ HL. Il poro MspA ha un'unica zona di riconoscimento dove i nucleotidi influenzano la corrente ionica, mentre l' $\alpha$ HL ne presenta varie (circa 12 nucleotidi).

ssDNA omo-polimerico viene ben discriminato dalla corrente di ostruzione, è possibile anche distinguere poli-mC da poli-dC. La discriminazione ha un'alta risoluzione anche sul singolo nucleotide.

In conclusione la ristrettezza della singola zona di interazione del MspA gli conferisce

una buona sensibilità al tipo di base e una risoluzione nitida della posizione del nucleotide. Queste caratteristiche ne fanno un ottimo candidato come poro per il rapido e preciso sequenziamento del DNA.

# Capitolo 4

## Conclusioni

In conclusione possiamo affermare che il sequenziamento del DNA, e dei polimeri in genere, tramite nanopori è un metodo efficiente ed efficace. Ha già dimostrato molte buone caratteristiche ma anche molte potenzialità ancora da sviluppare a pieno. Con questo metodo è possibile sequenziare in tempi brevi, filamenti molto lunghi, migliaia di paia di basi per iterazione. Questo numero non è limitato dal processo in sè, ma dalla preparazione e dalla qualità del campione. Ma già allo stato attuale, è molto competitivo rispetto ai metodi tradizionali come il Sanger, molto più lento, costoso e limitato nella lunghezza di lettura, e anche ad altri metodi di nuova generazione, non necessita di tanti reagenti quanto il pirosequenziamento o il sequenziamento a colorante, ma come questi può essere altamente parallelizzato.

Attualmente i nanopori e l'interazione nanoporo-polimero, presentano caratteristiche che ancora non comprendiamo o non sappiamo quantificare, il che rende questa tecnica un campo di ricerca in espansione. Il percorso per arrivare alla loro comprensione, pone nuove sfide alla nostra capacità di manipolare e studiare processi fisici, che coinvolgono solidi, liquidi e biomolecole, alla scala nanometrica.

Nonostante i buoni risultati già ottenuti, sono ancora molti i problemi da risolvere. La risoluzione va migliorata, in quanto per sequenziare con precisione un campione la misura va effettuata più volte (anche se parallelizzando il processo per più copie il problema si riduce), alcune tecniche devono ancora essere approfondite e perfezionate (electron tunneling).

# Bibliografia

- [1] David L. Nelson, Michael M. Cox: *I principi di biochimica di Lehninger*, Zanichelli, 2014
- [2] Michael Zwolak and Massimiliano Di Ventra: *Colloquium: Physical approaches to DNA sequencing and detection*, Reviews of Modern Physics, Volume 80, Gennaio-Marzo 2008.
- [3] Michael Zwolak and Massimiliano Di Ventra: *DNA Sequencing via Electron Tunneling*, Department of Physics Oregon State University, Department of Physics University of California, San Diego, 2012.
- [4] David Stoddart, Giovanni Maglia, Ellina Mikhailova, Andrew J. Heron, and Hagan Bayley: *Multiple Base-Recognition Sites in a Biological Nanopore: Two Heads are Better than One*, Angewandte Chemie International Edition, Volume 49, 2010.
- [5] David Stoddart, Andrew J. Heron, Ellina Mikhailova, Giovanni Maglia, and Hagan Bayley: *Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore*, Department of Chemistry, University of Oxford, United Kingdom, Daniel Branton, Harvard University, Cambridge, 2009.
- [6] Elizabeth A. Manrao, Ian M. Derrington, Mikhail Pavlenok, Michael Niederweis, Jens H. Gundlach: *Nucleotide Discrimination with DNA Immobilized in the MspA Nanopore*, Jerome Mathe, Université d'Evry val d'Essonne, France, 2011.
- [7] James Clarke, Hai-Chen Wu, Lakmal Jayasinghe, Alpesh Patel, Stuart Reid and Hagan Bayley: *Continuous base identification for single-molecule nanopore DNA sequencing*, Nature Nanotechnology Vol. 4, Aprile 2009.

- [8] Mingsheng Xu, Daisuke Fujita, and Nobutaka Hanagata: *Perspectives and Challenges of Emerging Single-Molecule DNA Sequencing Technologies*, Small, Vol. 5, 2009.
- [9] Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, et al.: *The potential and challenges of nanopore sequencing*, Nature Biotechnology, Vol. 6, 2008.
- [10] Michael L. Metzker: *Emerging technologies in DNA sequencing*, Genome Research, 2005.
- [11] Bruno Canard and Robert S. Sarfati: *DNA polymerase fluorescent substrates with reversible 3'-tags*, Gene, 1994.
- [12] Oxford Nanopore Technologies: [nanoporetech.com](http://nanoporetech.com)