

## COMPUTER A DNA INTRODUZIONE

I computer a DNA sono computer realizzati attraverso composti chimici al posto del software e DNA al posto dei dischi fissi. In pratica il DNA viene usato per archiviare i dati e per risolvere le operazioni del computer.

Utilizzando il DNA come software e gli enzimi come hardware, il computer a DNA è estremamente potente. Nel 2002, dei ricercatori in Israele hanno creato un computer a DNA che potrebbe eseguire 330 miliardi di operazioni al secondo, più di 100.000 volte più veloce rispetto alla velocità del PC più veloce.

Inoltre grazie a questa tecnica stanno nascendo i computer molecolari che utilizzano singoli atomi o molecole come mezzo per risolvere problemi computazionali. Tutte le forme di calcolo molecolare sono attualmente in fase iniziale, ma si prevede che presto sostituiranno i computer tradizionali.

### ***Esempio :***

Immaginate un computer che ricerca l'uscita da un labirinto e arriva a un punto critico, con un percorso che porta a sinistra e l'altro a destra.

I calcolatori elettronici devono scegliere quale percorso seguire per primo, mentre il nuovo computer non ha bisogno di scegliere, perché può replicare sé stesso e seguire entrambi i percorsi allo stesso tempo, trovando la risposta più rapidamente.

Tutto ciò è possibile perché i processori sono fatti di DNA anziché di silicio. Tutti i computer elettronici hanno un numero fisso di chip.

L'abilità del nostro computer di crescere, attraverso la duplicazione del DNA, mentre svolge calcoli lo rende più veloce di qualsiasi altra forma di computer e gli permette di trovare soluzioni per molti problemi considerati precedentemente impossibili.

### ***L'esperimento di Adleman***

Il primo esperimento realizzato attraverso il DNA computing avvenne nel novembre del 1994 e fu realizzato da Adleman.

Consisteva nel risolvere un problema di cammini hamiltoniani (in un grafo trovare quel percorso che tocca tutti i nodi e ogni nodo deve essere toccato solo una volta) e grazie all'uso del DNA come codifica per il grafo il problema è stato risolto.

Infatti ogni base del DNA (adenina, citosina, guanina, timina) codifica 2 bits (0,1) e pertanto ogni grammo di DNA codifica al massimo  $2 \times 10^{21}$  bits mentre con le tecnologie attuali non si possono immagazzinare più di  $10^9$  bits per grammo.

## ARCHIVIAZIONE DATI NEL DNA

Ogni base del DNA, cioè Citosina, Guanina, Timina e Adenina codificano due bits a testa. Secondo questo tipo di codifica:

Two bits	Nucleotides
00	A (Adenine)
10	G (Guanine)
01	C (Cytosine)
11	T (Thymine)

In questo modo, siamo in grado di utilizzare le 4 basi che compongono il filamento di DNA per codificare ogni byte di dati.

Grazie a questa codifica si possono archiviare grande quantità di dati nel DNA e soprattutto trasformare dati in binario in basi del DNA.

**Esempio:** Vogliamo salvare la parola ciao nel DNA.  
Per prima cosa trasformiamo ciao in binario :

**ciao = 01100011 01101001 01100001 01101111**

Adesso codifichiamo ogni coppia di bit nella rispettiva base di DNA:

**01 = C**  
**10 = G**  
**00 = A**  
**11 = T**

E così per tutto il resto del numero binario  
Il valore risultante sarà:

**CGAT CGGC CGAC CGTT**

Perciò quando poi successivamente si analizzerà il DNA, in particolare questo filamento si potrà decodificare leggendo i valori delle basi ed eseguendo i passaggi a ritroso della codifica.

Inoltre grazie alle caratteristiche del DNA questo dato potrà essere salvato in modo permanente e rimanere invariato per migliaia di anni, infatti il DNA può contenere informazioni per migliaia di anni senza che i dati si deteriorino.

Inoltre grazie a questa tecnica si possono salvare qualsiasi tipo di file: audio, immagini e testo.

Infatti dei ricercatori della Microsoft, utilizzando metodi di creazione artificiale del DNA, sono riusciti a codificare i bit di un file di 200 MB in una catena di molecole, in particolare adenina, timina, guanina e citosina, appunto le unità base del DNA. Tra i dati salvati nel minuscolo filamento anche il video di *This Too Shall Pass* degli OK Go in qualità HD.

Rimanendo nel fatto che quel filamento che codifica la stringa ciao è minuscolo e va ricordato che ogni grammo di DNA codifica al massimo  $2 \times 10^{21}$  bits mentre con le tecnologie attuali non si possono immagazzinare più di  $10^9$  bits per grammo.

## ELABORAZIONE

Nel primo esperimento del computer DNA, che fu un'importante prova di elaborazione tramite DNA, il bio-informatico Adleman riuscì a risolvere il Problema del percorso hamiltoniano per un grafo orientato (PPHO), che chiede di stabilire se in un dato grafo orientato in cui siano scelti due vertici, esiste almeno un cammino hamiltoniano che li unisce.

Un cammino hamiltoniano è quel percorso in un grafo orientato che tocca solo una volta tutti i vertici del grafo.

Adleman dopo una settimana di lavoro in laboratorio riuscì ad arrivare alla soluzione e alla creazione di una nuova tecnica di calcolo.

Per descrivere il procedimento di Adleman prendiamo come riferimento un grafo con quattro vertici X, Y, Z, W e un cammino di sei lati, nel quale si vuole determinare l'esistenza di un percorso hamiltoniano che inizia nel vertice X e termina nel vertice Z. Inoltre codifichiamo i vertici con sequenze di soli 8 basi di DNA. Per motivi che saranno chiariti in seguito accanto al codice di ogni vertice scriviamo anche il suo complemento ricordando che le coppie di basi complementari sono, {A, T}, {G, C}.

In questo modo:

Vertice	Codice DNA	Complemento
X	ACTTGCAG	TGAACGTC
Y	TCGGACTG	AGCCTGAC
Z	CCGAGCAA	GGCTCGTT
W	GGCTATGT	CCGATACA

Adesso codifichiamo ogni lato con una sequenza sempre di 8 basi di DNA, ma a differenza dei vertici che erano codificati con sequenze casuali i lati hanno una regola per essere codificati.

Infatti ogni sequenza è formata con la seconda metà della sequenza che identifica il primo vertice e la prima metà di quella del secondo vertice. Se prendiamo la codifica del vertice X che è ACTTGCAG e il vertice Y che è TCGGACTG per formare il lato XY dobbiamo prendere gli ultimi 4 valori di X perciò GCAG e i primi 4 di Y cioè TCGG, infine risulterà che il lato XY è GCAGTCGG.

E così per tutti i lati.

Lati	Codice DNA
XY	GCAGTCGG
XZ	GCAGCCGA
YW	ACTGGGCT
YZ	ACTGCCGA
YX	ACTGACTT
WZ	ATGTCCGA

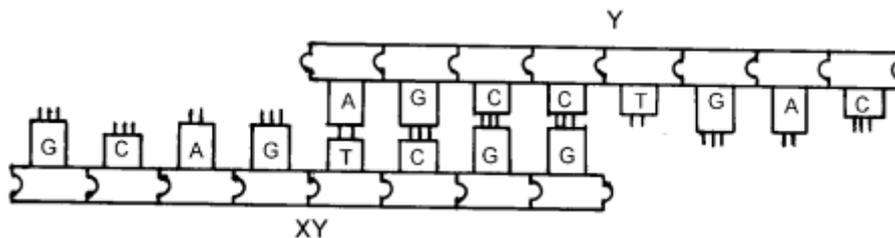
Codificati i nomi dei vertici e dei lati si sintetizzano filamenti di DNA corrispondenti ai codici dei lati e ai complementari dei codici dei vertici.

Il processo viene poi realizzato mettendo in una provetta un numero sufficientemente grande di filamenti di ciascuno dei DNA sintetizzati assieme ad acqua, enzimi, sali ed altre sostanze che favoriscano il processo biologico dell'accoppiamento. In questo modo si formano istantaneamente (quasi) tutti i possibili cammini tra i vertici.

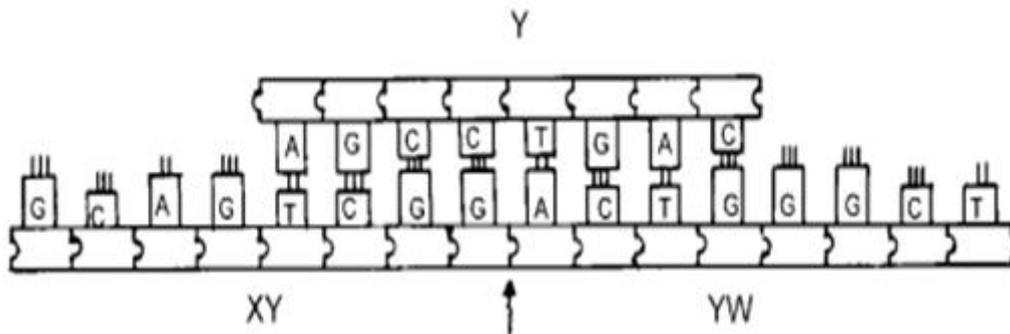
Il modo in cui i lati si uniscono per formare il percorso è questo:

Quando per esempio il lato XY, GCAGTCGG, si vuole collegare con il lato YW, ACTGGGCT, bisogna usare il complementare del vertice Y, AGCCTGAC (dato che è il vertice in comune ai due lati).

Questi due lati si incontrano, poiché la seconda parte di XY, TCGG, è complementare alla prima parte del complemento di Y, AGCC, e le due estremità si collegano.



Inoltre la prima parte di YW, ACTG, si collega con la seconda parte del complementare, TGAC.



In questo modo si formano catene che codificano i lati e sono tenute insieme dai DNA complementari dei vertici. Nella provetta quindi si formano molecole di DNA che codificano percorsi casuali attraverso i vertici del grafo, così come richiesto. Tenuto conto del numero assai grande di molecole messe in gioco è possibile che si siano formati (quasi) tutti i cammini possibili e che quindi almeno una delle molecole presenti nella provetta codifichi l'eventuale percorso hamiltoniano.

La provetta dopo questa prima operazione contiene la risoluzione, i passi successivi sono volti alla sua identificazione, cioè alla sua estrazione dalla massa di molecole codificanti cammini, che si sono formate assieme a lei.

Successivamente bisogna scartare tutti i percorsi che non iniziano e terminano nei vertici desiderati e separare tutte le molecole di DNA che hanno la lunghezza giusta: 24 nel caso del nostro esempio, dove il cammino hamiltoniano deve risultare lungo 3 (numero di nodi - 1) ed ogni lato è codificato da 8 lettere.

Infine bisogna eliminare i percorsi che non passano per tutti i vertici, sfruttando il principio di complementarità delle basi, e mediante il quale partendo dalla codifica di un vertice si separano tutte le sequenze che contengono quel vertice. È chiaro che dopo aver effettuato il procedimento per ogni vertice del grafo, le eventuali molecole di DNA che rimangono passano per tutti i vertici una ed una sola volta.

Questo è il passo che sperimentalmente risulta più lungo e complicato.

Nel caso del nostro esempio si dovrebbe ottenere la sequenza GCAGTCGGACTGGGCTATGTCCGA, che codifica il cammino (XY)(YW)(WZ).

Alla conclusione dell'esperimento lo scopo di Adleman non era tanto quello di risolvere il problema di Hamilton, bensì quello di dimostrare la possibilità di calcolare con il DNA.

## VANTAGGI DELL' USO DEL DNA

- Capacità enorme
  - Il computer a DNA immagazzina per ogni filamento  $10^{21}$  bit, in più lavora in parallelo quindi il risultato ottenuto si deve raddoppiare.
- Fedeltà di transizione
  - Il computer a DNA riesce a sostenere  $10^9$  transizioni al secondo con una fedeltà di transizione superiore al 99,8%
- Consumo di elettricità
  - Il computer a DNA per riuscire a funzionare consuma una quantità infinitesimale di elettricità:  $10^{-10}$  W.

## SVANTAGGI DELL'USO DEL DNA

- Costi molto alti
  - Infatti archiviare dati nel DNA a livello concreto è molto costoso in quanto prevede l'uso di molte elaborazioni chimiche che si possono eseguire solo in laboratorio e con strumenti chimici. Infatti in un esperimento del 2017 della Columbia University i costi per memorizzare 2 MB erano di 7000 \$ e per leggerlo 2000 \$. Anche per questo motivo adesso ci si limita solo ad esperimenti e non ad un uso più regolato e frequente.
- Tecniche molto complesse
  - Infatti le tecniche usate per archiviare i dati e modificarli sono molto complesse e vengono eseguite solo in laboratorio e spesso si realizzano degli errori durante queste operazioni.

### *Dna VS computer elettronici*

- Allo stato attuale, gli algoritmi del computer a DNA non son in concorrenza con gli algoritmi all'avanguardia dei calcolatori elettronici
- Procedure di laboratorio che richiedono tempo e tecniche complesse mentre i computer attuali sono semplici da usare e utilizzabili da tutti
- Nessun metodo universale di rappresentazione dei dati, mentre nei computer elettronici è usato da tutti il sistema binario

## BIBLIOGRAFIA

- Descrizione e introduzione al computer DNA :
  - <https://www.nature.com/articles/35106533>
- Esperimento Microsoft codifica video sul DNA :
  - <https://www.smartworld.it/informatica/microsoft-200-mb-dna.html>
- Esperimento per salvare file su DNA :
  - <https://www.ninjamarketing.it/2016/08/04/salvare-dati-digitali-sul-dna-da-oggi-di-puo/>
- Codifica da DNA a binario :
  - <https://mitjafelicijan.com/encoding-binary-data-into-dna-sequence#encode-binary-data-into-dna-sequence>
- Fase di elaborazione di un computer DNA :
  - <http://www.dipmat.univpm.it/~montecchiari/infstat/materiale/calc-DNA.htm>