

Modellizzazione del DNA con automi cellulari: simulazione del comportamento strutturale e dinamico di sequenze

Elena *Calcagni*^{*}, Katia De *Arcangelis*^{*},
Ernesto Di *Mauro*⁺, Patrizia *Mentrasti*^{*}

Dipartimento di *Matematica*^{*} e
Dipartimento di Genetica e Biologia *molecolare*⁺
Università degli Studi di Roma "La Sapienza"

P.le Aldo Moro, 2, 00185 Roma

Email: calcagni@mat.uniroma1.it, ernesto.dimauro@mat.uniroma1.it,
mentrasti@mat.uniroma1.it

Abstract

Descriviamo tre diversi modelli di simulazione della dinamica di sequenze di DNA, immerse in un fluido non neutro a temperatura costante, insieme all'implementazione realizzata in ambiente software SIMP/STEP, dedicato alla simulazione di automi cellulari. Prendendo in considerazione singoli filamenti di doppia elica e imponendo che vengano rispettate le leggi della fisica classica e condizioni aggiuntive che chiamiamo di "connettività" e "volume escluso", attribuiamo ai monomeri velocità di movimento diverse e specifiche, determinate in relazione alla temperatura media di fusione di coppie di basi in cis. L'implementazione ha permesso di effettuare misurazioni del movimento del baricentro delle stringhe nel caso di figure chiuse (rombo) e di figure aperte (rombo aperto); confrontando tra loro i risultati ottenuti si può affermare che le estremità libere influiscono sul movimento e si può valutare quantitativamente quest'effetto. Inoltre, studiando le stringhe aperte determiniamo in che modo la qualità dei monomeri terminali influenzi il movimento globale della molecola. L'emergere di queste proprietà legate alle estremità richiede, quindi, una complessità minima del sistema. Al di sotto di una lunghezza minima la proprietà "effetto di estremità" non compare.

Keywords: automi cellulari, sequenze di DNA, simulazione, sistemi dinamici

1 Introduzione

Vogliamo qui analizzare le proprietà strutturali intrinseche del DNA che appaiono emergere dall'analisi del comportamento asintotico di modelli di rappresentazione della dinamica di stringhe¹. Il nostro approccio è fenomenologico ed è basato sulla creazione di programmi per computer che simulano il comportamento di insiemi di unità. Le caratteristiche di queste unità collegate in stringhe sono programmate per simulare, in modo semplificato e modellizzato, le proprietà di base del DNA.

L'idea di base consiste nel simulare il comportamento dei singoli componenti adottando le seguenti scelte:

- tenendo conto della struttura primaria del DNA² differenziamo tra loro i i monomeri [2];
- basandoci sulla struttura secondaria³ [3] e utilizzando i risultati di una analisi di Gotoh e Tagashira [4] in cui viene descritta la variazione della temperatura di fusione locale della struttura a doppia elica [5], moduliamo la velocità di reazione delle basi cioè il movimento nello spazio;
- trascuriamo la struttura terziaria [6] in quanto la nostra simulazione è in due dimensioni.

Rimandiamo la paragrafo 2 per la descrizione tecnica e più dettagliata del nostro strumento, gli automi cellulari; ci limitiamo qui a dire che pensiamo le stringhe immerse in un fluido e immaginiamo di osservarne comportamento, imponendo loro di muoversi in uno spazio a due dimensioni e di rispettare le usuali regole della fisica classica. In effetti, nei nostri modelli che sono sulla linea di [21] rappresentiamo i monomeri come "palline" sottoposte alle leggi di conservazione della massa e della quantità di moto e ad alcune regole specifiche [8] che ne governano il movimento: in particolare una regola

¹in teoria dei linguaggi formali [1], una stringa è una sequenza finita di caratteri (o simboli) da un insieme finito detto alfabeto. Nel nostro caso l'alfabeto è rappresentato da {A, T, G, C}

²Per sequenza primaria del DNA si intende il susseguirsi lineare delle strutture unitarie componenti. Le strutture unitarie sono in questo caso i nucleotidi, molecole composte di una parte invariante -deossiribosio legato ad un residuo di fosfato- e di una parte variabile, costituita da una delle quattro basi azotate, Timina (T), Adenina (A), Guanina (G), Citosina (C). Il DNA è chimicamente e strutturalmente polare; esso è quindi caratterizzato da un inizio ed una fine e, nel caso di molecole chiuse, da una direzione chimica.

³la struttura a doppio filamento chimicamente complementare e direzionalmente antiparallela: la doppia elica.

non permette ai monomeri di avvicinarsi oltre la distanza del raggio del monomero, e un' altra non permette loro di rompere i legami della catena, cioè di allontanarsi oltre il punto di rottura.

Inoltre rappresentiamo la doppia elica come serie di monomeri intrinsecamente collegati (la base A con la base T, e la base G con la C), ma escludiamo la possibilità di separazione tra i due filamenti.

Il movimento di "filamenti di DNA" è stabilito determinando la velocità dei monomeri in relazione alle condizioni chimico-fisiche dell'ambiente in cui si trovano: l'effetto di queste condizioni è direttamente valutabile dalla misura della destabilizzazione della struttura della doppia elica e quest'ultima viene correlata quantitativamente con la temperatura media di fusione della doppia elica, cioè con la temperatura che mediamente è necessaria per dividere la doppia elica in due filamenti: risultati sperimentali[4] stabiliscono che quanto più tale temperatura è alta, tanto più le molecole si muovono lentamente nella reazione all'ambiente circostante.

Il nostro studio strutturale si basa sull' osservazione del comportamento dinamico dei filamenti. A questo scopo, selezioniamo per lo studio alcune variabili globali che possano rivelare l' emergenza di proprietà strutturali derivanti non dal comportamento delle singole componenti, ma dalla loro interazione e dal fenomeno osservato nella sua interezza.

Le variabili che abbiamo scelto sono la distanza tra le due estremità per stringhe aperte di diversa lunghezza [10] e il movimento del baricentro del sistema nel caso di figure chiuse e aperte [11] studiandone il comportamento nel tempo.

Quello che risulta emergere dalle nostre simulazioni sono le differenze tra:

- un sistema aperto e uno chiuso
- stringhe corte e stringhe lunghe
- un sistema isolato e uno affollato.

Gli strumenti che abbiamo utilizzato sono sistemi dinamici rappresentati con automi cellulari [12] e implementati nell'ambiente di lavoro SIMP/STEP [13].

I nostri sistemi algoritmici si basano su:

1. principi di simulazione in parallelo del movimento di stringhe dovuti a Bar-Yam [8, ?];
2. la relativa implementazione realizzata da Margolus in ambiente CAM-8 [14];

3. le possibilità e l'elasticità fornite dall' integrazione tra il linguaggio Python [15] e il sistema SIMP/STEP, realizzato per il sistema operativo Linux, integrazione ottimale tra software e hardware;
4. confronti e test con stringhe reali di DNA [16].

Rimandiamo al sito in costruzione per una descrizione dettagliata dei programmi e della loro filosofia implementativa; qui descriviamo i risultati più interessanti emersi da anni di studi e di osservazioni.

Nonostante il modello realizzato sia solo un prototipo schematizzato e semplificato, alcune delle caratteristiche e dei risultati emersi dalle simulazioni risultano particolarmente rivelatori, in quanto forniscono una possibile interpretazione a problematiche genetiche ancora oggi irrisolte quali: i rapporti tra struttura locale del DNA e struttura globale; influenza delle estremità ("telomeri") sulle proprietà fisiche globali della molecola di DNA e rapporto tra organizzazione topologica della molecola e struttura locale.

2 I nostri strumenti

2.1 Automi cellulari

L'automa cellulare risulta essere uno degli strumenti più efficaci per la rappresentazione del comportamento dinamico e microscopico di elementi in cui le interazioni più importanti da prendere in considerazione sono solo a corto raggio.

Non è un caso che Von Neumann abbia scelto di definire questo modello proprio per rappresentare la capacità di autoriproduzione dei viventi e trasportarla nelle "molecole" dei computer che sono gli automi a stati finiti [1, 17].

Le componenti essenziali di un automa cellulare sono costituite da: le *celle* che sono poste su una griglia regolare e possono assumere uno stato qualunque da un insieme finito Q di stati; l' *intorno di vicinato* di ciascuna cella (ogni cella è centro del proprio universo), e l'*uniformità* di tutte le componenti, cioè della griglia, delle celle e dell'intorno.

Per uno studio più approfondito dell'automa cellulare si veda [12, 18].

Definizione 1 Un automa cellulare è una terna (d, N, f) , dove:

- $d \in \mathbb{N}$, è la **dimensione dello spazio** in cui si lavora, nei nostri modelli $d=2$;

- N è un sottoinsieme finito di \mathbb{Z}^d , detto **indice di vicinato**, con $|N|=n+1$, cioè ogni cella, oltre se stessa, ha n vicini. Il vicinato rappresenta l'insieme dei punti dello spazio che influenzano direttamente ogni cella;
- f è la **funzione di transizione locale** che regola il comportamento delle celle, cioè modifica ognuna di esse in base allo stato della cella stessa e di quello delle vicine.

L'insieme delle celle forma quindi uno spazio d -dimensionale: ad ognuna di esse è associata una variabile locale di stato che può assumere valori finiti in un insieme detto insieme degli stati Q .

Una volta che siano definite (d, N, f) e una configurazione iniziale, l'automa cellulare può, in teoria, evolvere fino all'infinito.

Definizione 2 Una **configurazione** C , di un dato automa cellulare (d, N, f) , è descritta assegnando ad ognuna delle celle dell'automa al tempo t uno stato da un insieme di stati prefissato Q :

$$C : \mathbb{Z}^d \longrightarrow Q$$

Chiamiamo C_0 la configurazione iniziale al tempo $t=0$.

È interessante notare che la funzione di transizione f induce una funzione di transizione globale F , che mappa configurazioni in configurazioni. In base alle leggi locali, un compito molto arduo è individuare proprietà della F (invarianti) che sono determinate da f .

Una configurazione è un'assegnazione di stati a tutte le celle dell'automa cellulare; l'emergere di proprietà globali correlate con la funzione globale viene descritto dall'evoluzione di opportune variabili globali.

L'ambiente in cui sono stati implementati i modelli è il *SIMP/STEP* [13] è una macchina virtuale completamente software, disponibile per il sistema operativo Linux, in grado di supportare esperimenti che simulano automi cellulari anche molto complessi; il linguaggio di programmazione sul quale si basa l'interfaccia grafica è il python, un linguaggio di programmazione moderno, interpretato e di notevole potenza [15].

Rimandiamo al sito per domande e curiosità.

2.2 Parallelizzazione della dinamica di stringhe

La prima cosa che chiediamo ai modelli che saranno descritti è che possano essere adattabili ad algoritmi paralleli, in modo da poterli simulare più efficientemente.

Il nostro scopo è simulare il movimento di monomeri rappresentati da "palline" immerse in un fluido, alla temperatura di 30°; le "palline" si possono muovere nel fluido non casualmente, tenendo conto dei legami interni tra i monomeri.

Vogliamo quindi permettere il movimento nel nostro spazio di simulazione che è rappresentato da una griglia bidimensionale, 256×256 pixel tenendo saldi i legami intermolecolari.

Allo scopo di accelerare la velocità di movimento dei programmi di simulazione in parallelo, possiamo pensare di assegnare ad ogni monomero un processore diverso ed eseguire lo spostamento dei monomeri in modo locale. Potrebbe accadere, però, che un processore ad un passo non conosca la posizione corrente dei monomeri vicini e che quindi causi la separazione o la sovrapposizione di monomeri nella catena di DNA, andando quindi a modificare i legami interni.

Per ovviare al problema ciascun monomero ("pallina") deve soddisfare, oltre alle leggi della fisica classica [7], due importanti condizioni:

- (a) *connettività*, ogni monomero non si può staccare dai vicini;
- (b) *volume escluso*, ogni monomero non può sovrapporsi al volume di un altro.

Per soddisfare queste due limitazioni abbiamo adottato l'approccio originale di Bar-Yam [8] così da permettere in parallelo le simulazioni del movimento di metà dei monomeri ad ogni passo. Rimandando a [8] per una descrizione dettagliata dell'algoritmo e a [10, 11, 21] per la nostra implementazione, osserviamo che la nostra simulazione immagina le stringhe di polimeri poggiate su un piano (lo stesso per tutte le stringhe che evolvono insieme) e divise in posizioni di posti pari e dispari; in breve, ogni volta, scegliamo se far muovere i monomeri che si trovano nelle posizioni pari o dispari, e poi controlliamo le limitazioni imposte per metà dei monomeri presenti.

3 I nostri modelli

3.1 Differenziare i monomeri per peso molecolare

Il nostro primo modello vuole simulare la dinamiche dei "filamenti di DNA" distinguendo le basi dei monomeri a seconda del loro peso molecolare; lo

sviluppo di tale esperimento è già stato implementato con l'automa cellulare CAM-8 [19, 20, 21].

La differenziazione dei monomeri viene definita a seconda del peso molecolare delle basi del DNA; questa non avviene però tra le quattro basi distinte (A-T-G-C) e il relativo peso, ma considerando le basi accoppiate nella doppia elica.

Vengono distinti due tipi di monomero:

- 1) nucleotide con base A o T, che chiameremo di tipo AT (*trans*)⁴;
- 2) nucleotide con base G o C, di tipo GC (*trans*).

Rappresentiamo la doppia elica mediante un'unica catena e sottointendiamo la presenza della catena complementare, cioè considerando inscindibili i legami idrogeno A con T e G con C.

In questo modello, che chiamiamo *peso*, la differenziazione tra le basi di una catena di DNA viene definita non solo a seconda del peso molecolare di ciascuna base, ma per la diversa probabilità di movimento a seconda della direzione scelta per il movimento [22].

Base	Peso(in daltons)
Adenina (A)	135.13
Timina (T)	126.11
Guanina (G)	151.13
Citosina (C)	111.10

Nella doppia elica l'Adenina di un filamento si lega sempre con la Timina dell'altro (e viceversa), mentre la Citosina si lega sempre con la Guanina (e viceversa), possiamo quindi sommare i pesi delle basi che si legano ed ottenere dei valori non molto differenti tra loro:

$$p_{AT} = p_A + p_T = 135.13 + 126.11 = 261.24$$

$$p_{GC} = p_G + p_C = 151.13 + 111.10 = 262.23$$

Allo scopo di tener conto del peso atomico, stabiliamo che la probabilità di andare a nord e a sud sia proporzionale al peso atomico mentre la probabilità di andare ad est e ad ovest sia indipendente dal peso atomico e quindi uguale ad 1/4. Pertanto:

⁴con la notazione *cis* si indica il collegamento intramolecolare lungo lo stesso filamento del DNA; una base è legata in *cis* con quella precedente quella seguente attraverso legami covalenti. Con la notazione *trans* si indica il collegamento intermolecolare di una base con la base corrispondente sul filamento complementare di DNA; il legame in *trans* è costituito di legami idrogeno

- Le probabilità di movimento per l'Adenina sono:
 $\mathbb{P}(A \text{ vada ad } E) = \mathbb{P}(A \text{ vada ad } O) = 1/4$
 $\mathbb{P}(A \text{ vada a } N) = (1 - \mathbb{P}(A \text{ vada ad } E) - \mathbb{P}(A \text{ vada ad } O))p_T/p_{AT}$
 $\mathbb{P}(A \text{ vada a } S) = (1 - \mathbb{P}(A \text{ vada ad } E) - \mathbb{P}(A \text{ vada ad } O)) - \mathbb{P}(A \text{ vada a } N);$
- le probabilità di movimento per la Timina sono:
 $\mathbb{P}(T \text{ vada ad } E) = \mathbb{P}(T \text{ vada ad } O) = 1/4$
 $\mathbb{P}(T \text{ vada a } N) = (1 - \mathbb{P}(T \text{ vada ad } E) - \mathbb{P}(T \text{ vada ad } O))p_A/p_{AT}$
 $\mathbb{P}(T \text{ vada a } S) = (1 - \mathbb{P}(T \text{ vada ad } E) - \mathbb{P}(T \text{ vada ad } O)) - \mathbb{P}(T \text{ vada a } N);$
- le probabilità di movimento per la Guanina sono:
 $\mathbb{P}(G \text{ vada ad } E) = \mathbb{P}(G \text{ vada ad } O) = 1/4$
 $\mathbb{P}(G \text{ vada a } N) = (1 - \mathbb{P}(G \text{ vada ad } E) - \mathbb{P}(G \text{ vada ad } O))p_C/p_{GC}$
 $\mathbb{P}(G \text{ vada a } S) = (1 - \mathbb{P}(G \text{ vada ad } E) - \mathbb{P}(G \text{ vada ad } O)) - \mathbb{P}(G \text{ vada a } N);$
- le probabilità di movimento per la Citosina sono:
 $\mathbb{P}(C \text{ vada ad } E) = \mathbb{P}(C \text{ vada ad } O) = 1/4$
 $\mathbb{P}(C \text{ vada a } N) = (1 - \mathbb{P}(C \text{ vada ad } E) - \mathbb{P}(C \text{ vada ad } O))p_G/p_{GC}$
 $\mathbb{P}(C \text{ vada a } S) = (1 - \mathbb{P}(C \text{ vada ad } E) - \mathbb{P}(C \text{ vada ad } O)) - \mathbb{P}(C \text{ vada a } N).$

Tenendo conto dei legami idrogeno più forti tra le basi G e C, nel nostro modello poniamo una probabilità se pur piccola che il monomero G o C non si muova.

Le probabilità di movimento risulteranno quindi essere [21]:

- base Adenina (A):
 $\mathbb{P}(A \text{ si muove e va a } N) = 62/256$
 $\mathbb{P}(A \text{ si muove e va a } S) = 66/256$
 $\mathbb{P}(A \text{ si muove e va a } E) = \mathbb{P}(A \text{ si muove e va a } O) = 1/4$
 $\mathbb{P}(A \text{ non si muove}) = 0;$
- base Timina (T):
 $\mathbb{P}(T \text{ si muove e va a } N) = 66/256$
 $\mathbb{P}(T \text{ si muove e va a } S) = 62/256$
 $\mathbb{P}(T \text{ si muove e va a } E) = \mathbb{P}(T \text{ si muove e va a } O) = 1/4$
 $\mathbb{P}(T \text{ non si muove}) = 0;$
- base Guanina (G):
 $\mathbb{P}(G \text{ si muove e va a } N) = 54/256$
 $\mathbb{P}(G \text{ si muove e va a } S) = 73/256$
 $\mathbb{P}(G \text{ si muove e va a } E) = \mathbb{P}(G \text{ si muove e va a } O) = 1/4$
 $\mathbb{P}(G \text{ non si muove}) = 1/256;$

- base Citosina (C):
 $\mathbb{P}(C \text{ si muove e va a } N) = 73/256$
 $\mathbb{P}(C \text{ si muove e va a } S) = 54/256$
 $\mathbb{P}(C \text{ si muove e va a } E) = \mathbb{P}(C \text{ si muove e va a } O) = 1/4$
 $\mathbb{P}(C \text{ non si muove}) = 1/256.$

3.2 Assegnare il verso alla catena del DNA

Nella struttura secondaria del DNA, l'appaiamento delle basi A, T, G, C all'interno della doppia elica avviene solo tra basi complementari, ed è stabilizzato da legami idrogeno; in particolar modo le coppia G-C e C-G sono unite da tre legami idrogeno, mentre le coppie A-T e T-A da due legami.

La coppia G-C (C-G), proprio perchè unita da tre legami, si muove meno della coppia A-T (T-A), ma risultati ormai consolidati (modello di Gotoh e Tagashira [4]) hanno dimostrato che la differenza di movimento non dipende unicamente dal numero di legami tra le coppie, ma anche dalla base vicina secondo il verso della catena.

In particolare sono state calcolate sperimentalmente le temperature medie di fusione, T_m , di ogni singola coppia di basi complementari, in relazione alle diverse basi adiacenti.

È possibile schematizzare il valore delle 16 possibili coppie di basi, in relazione al vicino presente, tenendo conto che coppie complementari nel verso opposto di lettura della catena risulteranno avere, ovviamente, stessi valori di temperature medie di fusione:

	A	T	G	C
T	36.73°	54.50°	54.71°	86.44°
A	54.50°	57.02°	58.42°	97.73°
C	54.71°	58.42°	72.55°	85.97°
G	86.44°	97.73°	85.97°	136.12°

Nelle simulazioni supponiamo che il movimento avvenga in un fluido con temperatura media di 30°, in modo tale da essere sicuri che nessuna coppia arrivi alla temperatura di denaturazione e che ogni monomero rimanga quindi legato al nucleotide ad esso complementare (A-T, T-A, G-C e C-G).

Pertanto, assegnamo una probabilità di movimento tale che esso sia inversamente proporzionale ai valori di temperatura media di fusione, tenendo presente che le coppie con valori di temperatura di denaturazione più bassi si muovono di più rispetto a quelle con temperatura di denaturazione maggiore [10, 11].

Per assegnare le probabilità di movimento, attribuiamo alle coppie di basi dei monomeri con T_m più bassa la possibilità di muoversi sempre e calcoliamo i valori delle altre proporzionalmente, tenendo in considerazione che basi con temperatura media di fusione più alta si muovono di meno rispetto a quelle con temperature più basse. Stabiliamo una proporzionalità inversa tra temperature e possibilità di movimento.

La scelta della curva

$$y = \frac{k}{x}$$

dove con y indichiamo la probabilità di movimento, con x la temperatura media di fusione e con k la costante di proporzionalità inversa tra possibilità di movimento e temperatura media di fusione, comporterebbe valori per la probabilità di movimento troppo bassi: pertanto, dopo accurate verifiche sperimentali, abbiamo scelto di non utilizzarla.

Un buon compromesso è rappresentato dalla curva:

$$y = \frac{k}{(x + a)}$$

in cui stabiliamo che le coppie di basi con temperatura T_m minore, si muovano con probabilità del 100% e quelle con temperatura T_m più alta, si muovano con probabilità del 50%.

In conclusione, per differenziare le basi secondo il verso della catena, assegnamo nelle nostre configurazioni iniziali un valore diverso a ciascun monomero a seconda della base da cui è seguito, e conserviamo questo valore per tutta la durata della simulazione

Allo scopo di validare i nostri dati sperimentali abbiamo realizzato diversi modelli, raggruppiamo tra loro i valori di temperatura più vicini, in modo da utilizzare sono i valori medi, allo scopo di rendere più agevole l'esperimento.

EndToEnd

Nel primo modello *EndToEnd*, raggruppiamo le temperature ogni 30 gradi e di ciascun gruppo calcoliamo la media e la probabilità di movimento attribuendo alle coppie di basi dei monomeri con T_m più bassa la possibilità di muoversi sempre e calcolando i valori delle altre proporzionalmente:

	coppie di basi	media	Prob. movimento
(1)	TA, AA, TT, CA, TG, AT, CT, AG	52.28	1.00
(2)	CG, GG, CC, TC, GA, GT, AC	85.97	0.71
(3)	GC	136.12	0.50

Temp

In un secondo modello *temp*, raggruppiamo le coppie di basi con un intervallo di 20 gradi, calcoliamo la media per ciascun gruppo e la probabilità di movimento con proporzionalità inversa tra temperature e possibilità di movimento. :

	coppie di basi	media	Prob. movimento
(1)	TA, AA, TT, CA, TG	48.65	1.00
(2)	AT, CT, AG, CG	72.55	0.86
(3)	GG, CC, TC, GA, GT, AC	90.05	0.68
(4)	GC	136.12	0.50

Nei modelli *temp* e *EndToEnd*, si è definita all'interno delle configurazioni iniziali scritte in linguaggio Python una sequenza fissata (v6 in [16]):

CGG TG TTA CT CGGTG TAA CT

codificata in maniera opportuna per ciascun modello.

4 Le variabili di stato

4.1 Il baricentro

Definizione 3 Dato un sistema costituito da n punti materiali P_i ciascuno di massa m_i e vettore posizione $\vec{r}_i = (x_i, y_i)$, si definisce centro di massa o baricentro del sistema, il punto geometrico il cui vettore posizione \vec{r}_c è espresso dalla relazione:

$$\vec{r}_c = \frac{\sum m_i \vec{r}_i}{\sum m_i}$$

Abbiamo misurato nei nostri modelli lo spostamento del baricentro di diverse configurazioni iniziali; la parte principale del nostro studio verte sul confronto tra la dinamica del baricentro di figure chiuse (*rombo*), e quella di figure aperte (*rombo aperto*).

La nostra attenzione sullo studio delle configurazioni si basa sulla dinamica del movimento del baricentro di figure chiuse e aperte da cui emerge l'influenza dell'estremità nelle stringhe aperte; in particolare, abbiamo preso in considerazione il "rombo" e l'abbiamo posto a confronto con una figura

aperta che chiamiamo "rombo aperto", cioè con una figura che differisce dalla precedente per la mancanza di una sola base.

Dall'analisi dei nostri modelli è risultato che il movimento delle configurazioni risente molto della presenza o meno delle estremità [10]; gli estremi liberi tendono, infatti, a trascinare via il sistema consentendo movimenti più ampi e dimenticando prima la configurazione iniziale rispetto alle figure chiuse: non avendo nessun estremo libero hanno due gradi di libertà in meno.

Il modello *peso*, che differenzia i monomeri in base al peso molecolare delle coppie di basi unite da legami idrogeno, non attribuisce una stessa possibilità di movimento alle quattro direzioni possibili {nord, sud, est, ovest} e il baricentro delle figure è calcolato prendendo in considerazione il peso delle basi legate.

Il secondo modello *temp*, differenzia i monomeri in base al vicino presente secondo il verso della catena assumendo una divisione in quattro gruppi a seconda della temperatura media di fusione, attribuisce alle quattro direzioni possibili {nord, sud, est, ovest} la stessa possibilità di movimento (25%) e misura il baricentro delle figure considerando il peso delle basi uguale ad uno per ogni monomero.

Essendo questo sistema dotato di una simmetria interna, dovuta alle masse uguali per ogni base e alla stessa probabilità di andare nelle quattro direzioni, le diverse caratteristiche tra le figure aperte e chiuse vengono rilevate ad un numero di passi inferiore (quindi dopo minor tempo), rispetto ad un modello in cui abbiamo assenza di simmetria [11].

La scelta è stata quella di studiare il movimento di monomeri di una figura chiusa e quello di una figura aperta, facendo girare i nostri esperimenti per tempi piuttosto lunghi, 300000 passi temporali di simulazione e calcolando il baricentro della figura ad intervalli regolari.

Le configurazioni inserite nel modello sulle quali abbiamo eseguito le misure sono due:

- rombo con 146 elementi;
- rombo aperto con 145 elementi.

La distanza reale tra due monomeri nella molecola di DNA è 3.4 Å: nella nostra rappresentazione calcoliamo che 1 pixel = $3.39 \cdot 10^{-4}$ m = $3.39 \cdot 10^6$ Å. Se dividiamo la misura di un pixel per la distanza tra due monomeri ($3.39 \cdot 10^6 / 3.4$) otteniamo il fattore di conversione di 10^6 , cioè un pixel è 10^6 volte più grande della distanza tra due monomeri; quindi la scala da noi adottata nei modelli per una maggiore chiarezza nelle immagini è: $1:3 \cdot 10^{-6}$.

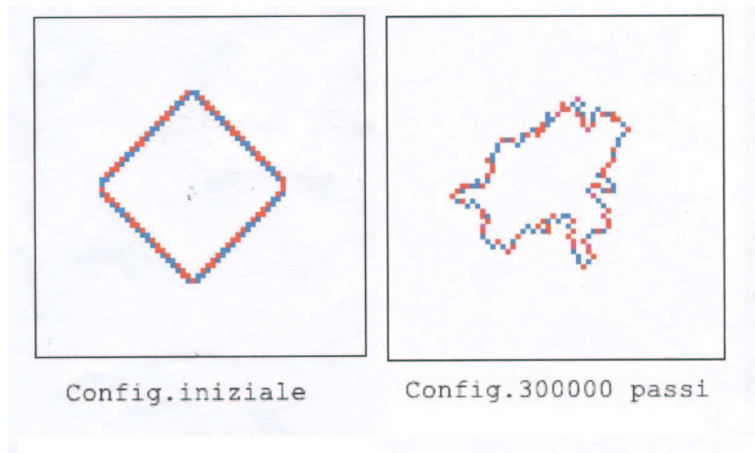


Figure 1: *Configurazioni rombo nel modello peso*

La prima configurazione rappresenta un rombo con 146 basi ottenute ripetendo la sequenza v6 di [16] più volte, centrata nel punto dello spazio di simulazione di $(x=128; y=128)$ (Fig.1, Fig.2).

La seconda è la figura di un rombo aperto con 145 basi ottenuta ripetendo la sequenza v6 di [16] più volte, eliminando però dalla precedente una base nel punto $(x=128; y=110)$, sempre centrata nel punto dello spazio di simulazione di $(x=128; y=128)$ (Fig.3, Fig.4).

Per poter studiare la dinamica del movimento del baricentro, abbiamo scelto di misurare la distanza euclidea d_1 che intercorre tra i punti del baricentro del rombo nei vari passi della simulazione (x_{bc}, y_{bc}) rispetto alla posizione iniziale del baricentro $(x_{bc0} = 128, y_{bc0} = 128.4998)$:

$$d_1 = \sqrt{(x_{bc} - x_{bc0})^2 + (y_{bc} - y_{bc0})^2}$$

la distanza d_2 tra i punti misurati nel caso del rombo aperto (x_{ba}, y_{ba}) sempre rispetto alla posizione iniziale $(x_{ba0} = 127.9999, y_{ba0} = 128.628)$

$$d_2 = \sqrt{(x_{ba} - x_{ba0})^2 + (y_{ba} - y_{ba0})^2}$$

e la distanza d_3 tra i punti del rombo (x_{bc}, y_{bc}) e del rombo aperto (x_{ba}, y_{ba}) ottenuti agli stessi passi di simulazione

$$d_3 = \sqrt{(x_{bc} - x_{ba})^2 + (y_{bc} - y_{ba})^2}$$

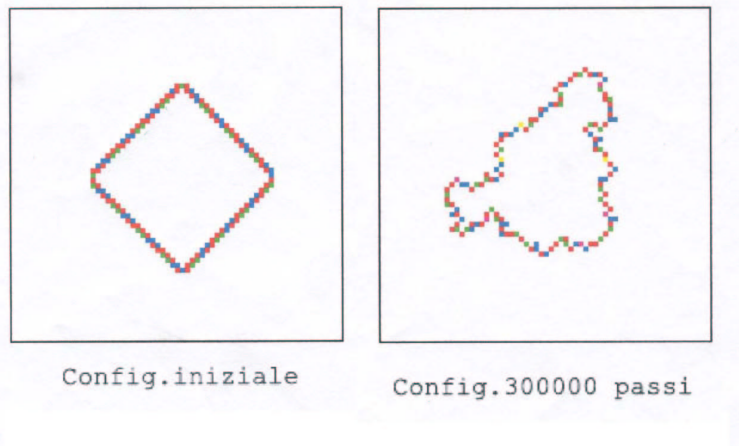


Figure 2: *Configurazioni rombo nel modello temp*

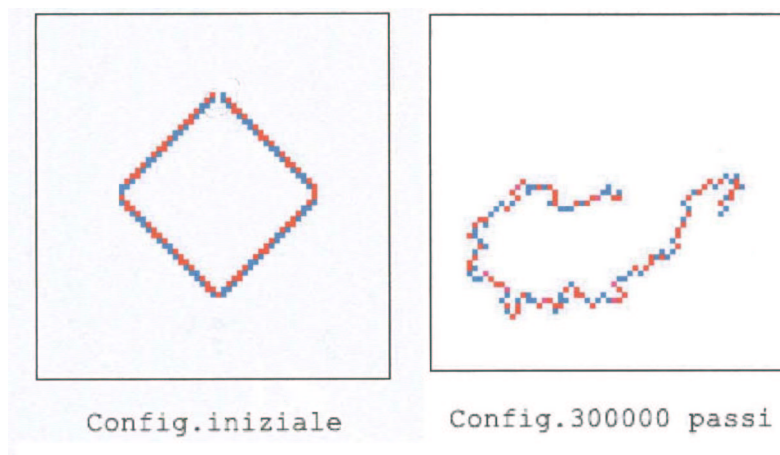


Figure 3: *Configurazioni rombo aperto nel modello peso*

4.2 La distanza tra le estremità

Definizione 4 *Siano (x_i, y_i) , (x_f, y_f) le coordinate dei monomeri terminali di una stringa polimerica, definiamo la distanza tra le estremità come:*

$$d = \sqrt{(x_f - x_i)^2 + (y_f - y_i)^2}$$

Le configurazioni inserite nel modello sulle quali abbiamo eseguito le misure sono due:

- stringhe con 146 elementi;
- stringhe con 20 elementi.

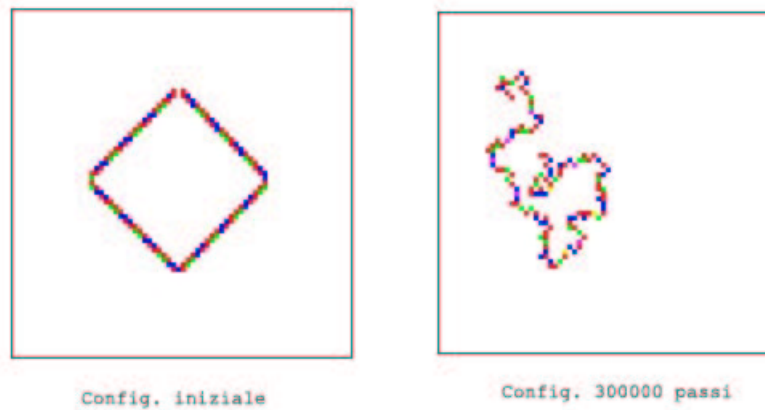


Figure 4: *Configurazioni rombo aperto nel modello temp*

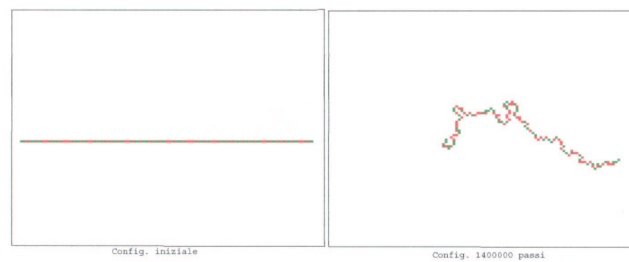


Figure 5: *Configurazioni della stringa1 nel modello EndToEnd*

Sono state effettuate un numero cospicuo di misurazioni di tre stringhe da 146 basi (Fig.5) , costituite dalla ripetizione dello stesso filamento ad eccezione dei due monomeri terminali. Abbiamo fatto tre diverse misurazioni con tre tipologie diverse di estremi:

- **stringa1** i cui monomeri terminali hanno la possibilità di muoversi massima del 100%;
- **stringa2** i cui monomeri terminali hanno la possibilità di muoversi media del 70%;
- **stringa3** i cui monomeri terminali hanno la possibilità di muoversi minima del 50%.

Diversi esperimenti effettuati sulla stessa tipologia di stringa hanno evidenziato comportamenti che tendono ad essere assimilati per tipologia. Infatti a partire da configurazioni iniziali identiche e a parità di passi :

1. si raggiunge uno stato pressochè stazionario per quanto riguarda la distanza tra le estremità; dopo un certo numero di passi, il valore della distanza comincia ad oscillare intorno ad un valore di "*persistence length (PL)*" ⁵;
2. il numero di passi, o tempo t_{PL} , necessario affinché la stringa raggiunga una definita PL è costante;
3. La lunghezza della traiettoria percorsa dalle estremità, L_e , si differenzia di poche unità.

La dinamica dei filamenti viene sensibilmente influenzata dalla tipologia delle basi terminali: infatti i tre valori PL , t_{PL} e L_e , che sono invarianti per filamenti identici, risultano essere diversi tra *stringa 1*, *stringa 2*, *stringa 3*.

Abbiamo fatto evolvere le tre diverse stringhe per 1400000 passi temporali di simulazione e calcolando i tre parametri ad intervalli regolari.

Abbiamo inoltre effettuato misurazioni anche su stringhe con 20 basi (Fig.6), al fine di verificare se proprietà osservate nelle stringhe con 146 basi fossero caratteristiche anche di stringhe con meno elementi.

- **20stringa1** i cui monomeri terminali hanno la possibilità di muoversi massima del 100%;
- **20stringa2** i cui monomeri terminali hanno la possibilità di muoversi media del 70%;
- **20stringa3** i cui monomeri terminali hanno la possibilità di muoversi minima del 50%.

In questo caso abbiamo fatto evolvere le tre diverse stringhe per un numero di passi inferiore al precedente, 700000 passi temporali di simulazione calcolando gli stessi tre parametri PL , t_{PL} e L_e ad intervalli regolari.

⁵la lunghezza di persistenza è una misura della lunghezza nella quale la catena "persiste". Essa è una media pesata della distanza end-to-end tra le estremità pensata per la lunghezza della catena che tende all'infinito.



Figure 6: Configurazioni della 20stringa1 nel modello EndToEnd

5 Proprietà emergenti

Studiare il movimento di "filamenti di DNA" è importante sia per approfondire la conoscenza delle proprietà interne della doppia elica, sia per capire come le diverse conformazioni da essi assunte influiscano sul cambiamento delle proprietà di interazione della molecola con l'esterno. In [23], ad esempio, viene illustrata un'ipotesi di come fluttuazioni locali nella struttura ad elica possano influenzare la lettura delle sequenze di basi di DNA da parte di proteine o enzimi di controllo [10].

5.1 Aperto-chiuso

Nel modello *peso* la figura del "rombo aperto", distinta da quella di "rombo" per la rimozione di una unità in posizione apicale, tende a muoversi e a cambiare la sua distanza dalla posizione iniziale del baricentro ($x=127.9999$; $y=128.6276$) in misura maggiore rispetto al "rombo" (sia in termini quantitativi assoluti che in termini cinetici). Dai grafici in appendice A (Graf 1) notiamo infatti che prima di poter parlare di situazione quasi stazionaria dei valori della distanza, devono passare circa 200000 passi di simulazione.

La rappresentazione grafica mostra che il baricentro del rombo aperto non rimane molto stabile rispetto alla sua posizione iniziale. L'ordinata del baricentro tende infatti ad aumentare ($127 < y < 146$) e l'ascissa, dopo una crescita iniziale, tende a diminuire ($100 < x < 132$). Il grafico (Graf 4) mostra comunque che il baricentro rimane piuttosto stabile nello spazio di simulazione tra le coordinate di 100-140.

Nella simulazione del rombo notiamo dal grafico in appendice A (Graf 1) che la distanza dalla posizione del baricentro iniziale ($x=128$; $y=128.4998$) cambia di poco e che, a differenza della figura aperta, quello che differisce maggiormente è il minor numero di passi necessario per arrivare ad una situazione quasi stazionaria. Un altro diverso comportamento tra i due sistemi è dovuto alla posizione del baricentro del rombo nello spazio di simulazione (Graf 3 in appendice A); notiamo infatti che il baricentro rimane piuttosto fermo ed oscilla tra le coordinate 120-140.

Osserviamo infine l'andamento del grafico in appendice A (Graf 2) della distanza tra il rombo e il rombo aperto: notiamo che il numero di passi (circa 120000) necessario per arrivare ad avere una situazione più o meno stazionaria dei valori della distanza tra le curve è dovuto al tempo necessario affinché vengano evidenziate le proprietà del sistema chiuso. In un primo momento la distanza tra le due curve è infatti minore perché l'andamento delle coordinate del rombo e del rombo aperto sono simili, mentre in un secondo momento differiscono fino ad arrivare ad una distanza di 28.

Analizzando la figura del rombo aperto nel modello per verso della catena *temp*, risulta che esso, avendo due estremità libere, tende a muoversi e a cambiare la sua distanza dalla posizione del baricentro iniziale ($x=128$; $y=128$) maggiormente rispetto al rombo, che raggiunge una posizione di quasi stazionarietà dopo circa 110000 come si nota dal grafico in appendice A (Graf 5).

Osservando il grafico in appendice A (Graf 8) nello spazio di simulazione, il baricentro pur variando non si discosta comunque troppo dalla posizione iniziale, rimanendo tra le coordinate di $100 < x < 130$ e $120 < y < 140$.

Nella simulazione del rombo notiamo dal grafico in appendice A (Graf 5) che la distanza dalla posizione del baricentro iniziale ($x=128$; $y=128$) cambia di poco e che non è facile parlare di posizione stazionaria in quanto non si hanno grandi discostamenti dalla posizione iniziale (128; 128).

Un importante comportamento diverso tra i due sistemi, aperto e chiuso, è dovuto alla posizione del baricentro del rombo nello spazio di simulazione (Graf 7 in appendice A); notiamo infatti che questo rimane fermo oscillando di una piccola quantità.

L'andamento del grafico (Graf 6 in appendice A) mostra la distanza tra il rombo e il rombo aperto, mettendo in luce che la distanza tra le due curve è inferiore rispetto al modello precedente e che questa varia maggiormente per il fatto che il rombo rimane quasi costante, mentre il rombo aperto si muove più liberamente in quanto non risente di vincoli cioè dell'estremità.

5.2 Simmetria e non-simmetria

L'asimmetria del sistema di *peso*, rispetto alla simmetria di *temp* serve per poter studiare la dinamica del nostro sistema: le proprietà intrinseche risultano le stesse, cambia però il tempo necessario per evidenziarle.

Il grafico in appendice A (Graf 9) mostra l'andamento della distanza tra figura aperta e chiusa, le curve fino a circa 100000 passi sono diverse in quanto in *peso* non c'è simmetria, mentre successivamente l'andamento tende ad essere più simile, anche se è diverso il tempo necessario per raggiungere una posizione quasi stazionaria.

5.3 Lungo-corto

Essendo costituite da un numero minore di basi, le stringhe più corte hanno una libertà di movimento molto maggiore; la lunghezza della traiettoria percorsa dalle estremità è, infatti, pressochè identica a quella che le molecole da 146 basi effettuano nel doppio del tempo.

Il sistema risulta coerente con le condizioni iniziali da noi imposte. Lo spazio L_e , percorso da ciascun tipo di estremità è infatti direttamente proporzionale alla proprietà di movimento loro assegnata.

La mobilità del sistema in questione non permette di definire una vera e propria *persistence length* comune a stringhe di tale lunghezza; calcoliamo la PL come media dei valori assunti dalla distanza end-to-end ogni 100000 passi, ma risulta che:

1. osservando i grafici in appendice A (Graf 10,11,12), indipendentemente dalle estremità considerate, non esiste un valore comune di PL ;
2. la PL calcolata più volte sulla stessa stringa può assumere valori che variano, rispetto alla lunghezza iniziale del filamento in un intervallo troppo ampio per poter descrivere uno stato stazionario.

I risultati emersi dalle simulazioni delle molecole con 20 basi, evidenziano, quindi, un comportamento che non permette di parlare di una reale influenza della tipologia dei monomeri terminali sul movimento globale del filamento; a

differenza delle stringhe da 146 basi, non abbiamo potuto quindi riscontrare delle proprietà caratterizzanti ciascuna stringa.

La qualità delle estremità gioca quindi un ruolo significativo in sistemi con maggiore stabilità, mentre ha un ruolo meno importante in sistemi meno vincolati e quindi più liberi. L'emergenza delle proprietà legate alle estremità richiede quindi una complessità minima. Al di sotto di questa lunghezza la proprietà "effetto di estremità" non compare.

6 Conclusioni

La molecola del DNA è estremamente complessa. La sua lunghezza può essere di decine di milioni di unità nucleotidiche, limitate soltanto da fattori legati ad efficienza di organizzazione genetica e di competizione replicativa. I meccanismi alla base di questa complessità strutturale sono di difficile studio e sono in genere analizzati su basi esclusivamente genetiche e chimico-fisiche. La modellizzazione bioinformatica è stata finora limitata dalla scarsità di strumenti adeguati e dalla mancata diffusione di modelli come gli automi cellulari [8, 9, 25]. Gli automi cellulari forniscono un approccio metodologico in grado di stabilire una base modellistica matematica. È in questo quadro di riferimento che il DNA viene considerato una "stringa" nell'accezione informatica di sequenza finita di elementi (caratteri) e che perde il suo significato genetico strutturale ma ne acquista uno informatico.

In questo tentativo di ricostruzione semantica, l'uso delle metodologie proprie degli automi cellulari ci ha permesso di ricreare in forma di modelli informatici le proprietà di base del DNA: la differenza di comportamento tra molecole lunghe e molecole corte, tra molecole chiuse (topologicamente definibili) e molecole aperte (che non lo sono), tra molecole asimmetriche e molecole dotate di simmetria interna. Il fatto che queste proprietà in qualche modo ovvie e di controllo sistemico siano state qui modellizzate pone le basi per lo studio di proprietà più complesse quali quelle topologiche e quelle di relazione intermolecolare. Il fatto che il metodo sviluppato abbia rivelato l'emergenza di proprietà legate alle estremità delle stringhe e ne abbia definito il presupposto di complessità è la prova della validità del metodo adottato.

Appendice A

Mostriamo in appendice i grafici riguardanti le proprietà emergenti dai nostri modelli:

- nel grafico 1. viene rappresentato l'andamento della distanza del baricentro dalla posizione iniziale del modello *peso*;
- nel grafico 2. l'andamento della distanza del baricentro tra il “rombo” e il “rombo aperto” del modello *peso*;
- nei grafici 3. e 4. viene rappresentato il baricentro nello spazio di simulazione del “rombo” (graf 3.) e del “rombo aperto” (graf 4.) del modello *peso*;
- nel grafico 5. viene rappresentato l'andamento della distanza del baricentro dalla posizione iniziale del modello *temp*;
- nel grafico 6. l'andamento della distanza del baricentro tra il “rombo” e il “rombo aperto” del modello *temp*;
- nei grafici 7. e 8. viene rappresentato il baricentro nello spazio di simulazione del “rombo” (graf 7.) e del “rombo aperto” (graf 8.) del modello *temp*;
- nel grafico 9. viene rappresentata la misura della distanza tra il baricentro del “rombo” e del “rombo aperto” tra i due modelli *peso* e *temp*;
- nei grafico 10. 11. e 12. viene rappresentato l'andamento della distanza end-to-end e la persistence length delle stringhe da 20 basi del modello *EndToEnd*;
- nei grafico 13. 14. e 15. viene rappresentato l'andamento della distanza end-to-end e la persistence length delle stringhe da 146 basi del modello *EndToEnd*.

Appendice B

Nell'ambiente **SIMP/STEP** il numero di bit messo a disposizione non è molto elevato, infatti si ha un numero massimo di 20 bit da suddividere tra tutti i piani necessari per la simulazione.

Nei modelli implementati sono stati definiti dei piani necessari all'implementazione [10] e [11]:

- **m0, mb0**: la combinazione dei quali sta ad indicare la presenza o l'assenza e il tipo del monomero sul *piano 0*.
- **m1, mb1**: la combinazione dei quali sta ad indicare la presenza o l'assenza e il tipo del monomero sul *piano 1*.
- **mk0, mk1**: usati per marcare i monomeri in cui si effettueranno le misure.
- **NORD, SUD, EST, OVEST**: contengono l'informazione sulla presenza (1) o assenza (0) dei vicini sul *piano 0*; il vicinato che si assume è quello di Von Neumann di raggio 1.
- **NORD0, SUD0, EST0, OVEST0**: contengono l'informazione dei vicini due celle più a nord, sud, est ed ovest sul *piano 0*; il vicinato che si assume è quello di Von Neumann di raggio 2.
- **rand0 e rand1** sono i piani random a due stati per assegnare la direzione del movimento: 0-0 NORD, 0-1 SUD, 1-0 EST e 1-1 OVEST.
- **randmov2, randmov3 e randmov4** sono i piani random a due stati per assegnare le probabilità di movimento.

Nelle simulazioni, le LUT, che regolano la dinamica delle stringhe polimeriche per il controllo delle regole di “*connettività*” e “*volume escluso*”, sono sostanzialmente quattro: *marca*, *dirMuovi*, *deposita* e *scambiaInfo*.

Negli algoritmi sono presenti altre funzioni, **faninfo**, **fanrand** e **fanout** che servono per inviare informazioni ai monomeri vicini o per ripulire i segnali per il passo successivo.

La funzione *marca*

La prima funzione *marca*, dopo che nella configurazione iniziale sono stati assegnati i monomeri sui due piani m0 ed m1 e i relativi tipi sui piani mb0 e

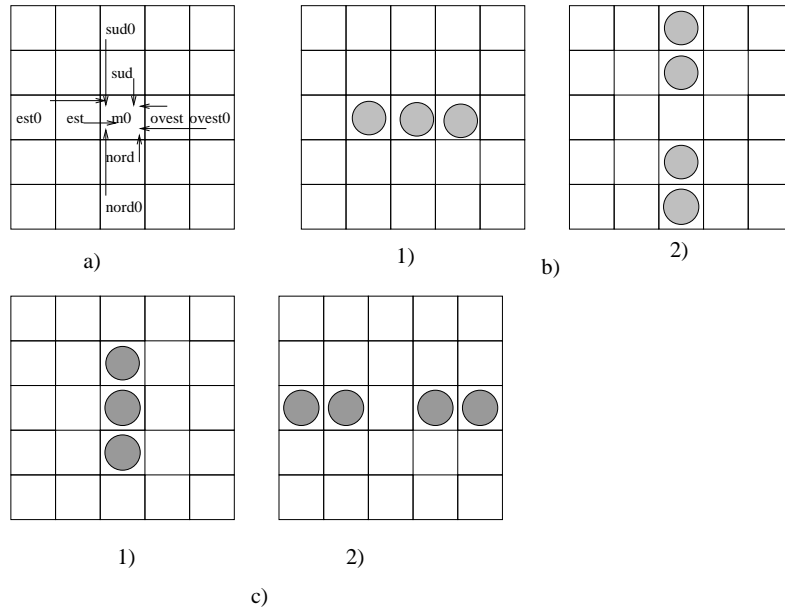


Figure 7: *Kick dei campi vicini alla cella centrale(a), controllo dei blocchi 1×3 (b) e 3×1 (c)*

mb1, esegue un test di presenza nei vicini necessari per la "connettività" e per il "volume escluso", ne copia il risultato in segnali d'appoggio che vengono poi calciati al posto giusto con un opportuno kick (Fig.7)(a).

Vediamo ora quello che accade osservando la Fig6.1(b)(1), ogni cella controlla se è presente almeno uno dei due vicini di est, ovest o se stesso e fa sapere il risultato ai vicini di nord, sud, nord0 e sud0 (Fig6.1(b)(2)) mentre nella Fig6.1(c)(1) il controllo sulla presenza dei vicini avviene per i blocchi 3×1 e il risultato è fatto conoscere alle celle orizzontali (Fig6.1(c)(2)).

La funzione *dirMuovi*

Nella LUT *dirMuovi* vengono verificate le condizioni di "connettività" e "volume escluso" che permettono lo spostamento di un monomero; la funzione verifica se il monomero è presente e se a quel passo di simulazione si può muovere controllando i piani randmov; a questo punto, se le due condizioni sono verificate, si sceglie casualmente, utilizzando i segnali rand0 e rand1, la direzione tra le quattro possibili con 25% di probabilità e poi, a seconda della direzione scelta, si controllano la "connettività" e il "volume escluso".

Se le verifiche per direzione scelta sono soddisfatte, si pone a zero il bit

m1 nel piano1 del monomero che deve essere spostato e vengono memorizzati il tipo e la posizione che dovrà assumere nella LUT successiva.

La funzione *deposita* e *ScambiaInfo*

Nella LUT *deposita* attraverso i piani di appoggio definiti è possibile posizionare i monomeri spostati nella nuova posizione: infatti ricordando tipo e direzione da assumere, memorizziamo in m1 ed in mb1 presenza e tipo e poniamo a zero i piani d'appoggio utilizzati. L'ultima funzione che regola il movimento dei monomeri è *scambiaInfo* e serve per invertire le informazioni presenti sul piano0 con quelle del piano1 in modo tale che al passo seguente siano diversi i monomeri di cui dobbiamo controllare il movimento e che eventualmente dobbiamo spostare.

References

- [1] Hopcroft J.E., Ullmann J.D., *Introduction to automata theory, languages and computation*, Addison Wesley, 1979
- [2] Wolfe S.L., *Introduzione alla biologia cellulare e molecolare*, EdiSES, Napoli, 1996
- [3] Watson J.D., *Biologia molecolare del gene*, Zanichelli, Bologna, 1972
- [4] Gotoh O., Tagashira Y., "Stabilities of Nearest-Neighbor Doublets in Double-Helical DNA Determined by Fitting Calculated Melting Profiles to Observe Profiles", *Biopolymers* (Vol.20, 1033-1042), 1981
- [5] Russell P.J., *Genetica*, EdiSES, Napoli, 1996
- [6] Mathews C.K., van Holde K.E, *Biochimica*, Ambrosiana C.E, Milano, 1994
- [7] Toffoli T., Margolus N., *Cellular Automata Machines - A new environment for modeling*, The MIT Press, Cambridge, Massachusetts, 1987
- [8] Bar-Yam Y., *Dinamics of Complex Systems (Studies in nonlinearity)*, Reading, Massachusetts, Addison-Wesley, 1997
- [9] Ostrovsky B., Crooks G., Smith M.A., Bar-Yam Y., "Cellular automata for polymer simulation with application to polymer melts and polymer

- collapse including implications for protein folding”, *Parallel Computing*, (27, 613-641), 2001
- [10] De Arcangelis K., *Automi cellulari per lo studio del comportamento strutturale e dinamico di filamenti di DNA*, Tesi di laurea in Matematica, Università di Roma "La Sapienza", a.a.2001-2002
- [11] Calcagni E., *Modelli di rappresentazione della dinamica di stringhe polimeriche con automi cellulari - Implicazioni per modelli genetici*, Tesi di laurea in Matematica, Università di Roma "La Sapienza", a.a.2001-2002
- [12] Spezzano G., Talia D., *Calcolo parallelo, automi cellulari e modelli per sistemi complessi*, F. Angeli, Milano, 1999
- [13] Franchini A., Mentrasti P., *Introduzione all'ambiente SIMP/STEP*, Dipartimento di Matematica "Guido Castelnuovo", Università degli Studi di Roma "La Sapienza", 2001/02
- [14] Mentrasti P., Tempestini O., *L'automa cellulare CAM-8: caratteristiche fisiche e dinamiche*, Dipartimento di Matematica "Guido Castelnuovo", Università degli Studi di Roma "La Sapienza" , 37/99
- [15] McDonald K., Harms D., *The Quick Python Book*, Manning, Greenwich CT, 2000
- [16] De Arcangelis V., Di Mauro E., Negri R., Travers A., Buttinelli M., Panetta G., “Sequence Dependence of Translational Positioning of Core Nucleosomes”, *J.Mol.Biol*(number 307, 987-999), 2001
- [17] Von Neumann J., *Theory of Self-Reproducing Automata*, edited and completed by Burks A., Univ. of Illinois Press, 1966
- [18] Clementi A., Mentrasti P., Pierini P., *L'automa cellulare, un modello di calcolo e di simulazione*, Dipartimento di Matematica "Guido Castelnuovo", Università degli Studi di Roma "La Sapienza", Novembre 1991
- [19] Tempestini O., Mentrasti P., *L'automa cellulare CAM-8: caratteristiche fisiche e dinamiche*, Dipartimento di Matematica "Guido Castelnuovo", Università degli Studi di Roma "La Sapienza" , 37/99
- [20] Tempestini O., Mentrasti P., *Istruzioni fondamentali per l'utilizzo della CAM-8*, Dipartimento di Matematica "Guido Castelnuovo", Università degli Studi di Roma "La Sapienza", 99/34

- [21] Astolfi S., *Simulazioni di dinamiche di stringhe di DNA con l'automa cellulare CAM-8*, Tesi di laurea, in Matematica, Università di Roma "La Sapienza", a.a.1999-2000
- [22] Astolfi S., Mentrasti P., *Simulation of DNA Strings Dynamics*, Dipartimento di Matematica "Guido Castelnuovo", Università degli Studi di Roma "La Sapienza", 18/2001
- [23] Dickerson R.E., *The DNA Helix and How It is Read*, Scientific American, Inc., Copyright 1995
- [24] Toffoli T., Bach T., *A common language for "programmable matter" (cellular automata and all that)*, Bulletin of the Italian Association for Artificial Intelligence, March 2001
- [25] Sirakoulis G.Ch., Karafyllidis I., Mizas Ch., Mardiris V., Thanailakis A., Tsalides Ph., "A cellular automaton model for study of DNA sequence evolution", *Computer in Biology and Medicine*, (33, 439-453), 2003