



SAPIENZA
UNIVERSITÀ DI ROMA

Studio tramite simulazioni "Coarse-Grained" al calcolatore della formazione di doppie eliche di DNA a partire da singoli nucleotidi.

Facoltà di Scienze Matematiche, Fisiche e Naturali
Laurea Magistrale in Fisica della Materia

Mattia Trapella
Matricola 1694399

Relatore
Prof. Cristiano De Michele

Anno Accademico 2022/2023

Studio tramite simulazioni "Coarse-Grained" al calcolatore della formazione di doppie eliche di DNA a partire da singoli nucleotidi.

Tesi di Laurea Magistrale. Sapienza Università di Roma

© 2023 Mattia Trapella. Tutti i diritti riservati

Questa tesi è stata composta con L^AT_EX e la classe Sapthesis.

Email dell'autore: trapella.1694399@studenti.uniroma1.it // mattia.trapella@gmail.com

Indice

Introduzione	1
1 Cristalli liquidi	5
1.1 Scoperta e sviluppo	5
1.1.1 Nomenclatura dei cristalli liquidi	7
1.2 Fasi liquido cristalline	7
1.2.1 Birifrangenza e analisi sperimentale dei cristalli liquidi	8
1.2.2 Fase solida e fase liquida	9
1.2.3 Fase isotropa	10
1.3 Fase nematica	10
1.3.1 Parametro d'ordine nematico	11
1.3.2 Fase colesterica o nematica chirale N^*	14
1.4 Fase Smettica	15
1.4.1 Fase Smettica A	15
1.4.2 Fase Smettica B	16
1.4.3 Fase Smettica C	16
1.5 Fase Colonnare Col	17
1.5.1 Parametro d'ordine esagonale	17
2 Parentesi Biofisica: DNA	19
2.1 Struttura chimica degli acidi nucleici	19
2.2 Struttura fisica del DNA	22
2.3 Parametri strutturali del DNA	23
2.4 Stabilità della doppia elica di DNA in soluzione	25
2.5 Forze di Pairing e Stacking	26
2.6 Cristalli liquidi di DNA	28
2.6.1 Filamenti di DNA lunghi	29
2.6.2 Filamenti di DNA corti	32
3 Simulazioni Monte Carlo	35
3.1 Importance Sampling	36
3.2 Il metodo Metropolis	38
3.2.1 Catene di Markov	38
3.2.2 Bilancio dettagliato	39
3.2.3 Codice	40
3.3 Ensembles	41

3.4	NVT	41
3.4.1	Implementazione codice	42
3.5	NPT	43
3.5.1	Basi di Meccanica Statistica	43
3.5.2	Implementazione Codice	45
4	Metodi: Approccio computazionale e design delle simulazioni	47
4.1	Presentazione dei due modelli	47
4.1.1	Cilindro	48
4.1.2	Cilindro con 2 e 4 patch	48
4.1.3	Semidisco con 4 patch	50
4.1.4	Semidisco con 4 patch e rotazione del filamento	51
4.2	Efficienza delle mosse	53
4.3	Periodic boundary conditions	53
4.4	Random Number Generator RNG	55
4.5	Generazione delle configurazioni iniziali	56
4.5.1	Configurazione iniziale cristallina	56
4.6	Impostazioni e ottimizzazioni del codice	57
4.7	Collisioni tra poliedri - SAT	57
4.7.1	Separating Axis Theorem	58
4.7.2	Proiezione su una linea	59
4.7.3	Poligoni convessi stazionari in 2D	59
4.7.4	Poligoni convessi stazionari in 3D	62
4.7.5	Implementazione e ottimizzazione codice	64
4.8	Collisione tra poliedri - <i>Xenocollide</i>	66
4.8.1	Mappatura di supporto	66
4.8.2	Differenza di Minkowski	67
4.8.3	Rilevamento di collisioni - 2D <i>Minkowski Portal Refinement</i>	68
4.8.4	MPR in 3D	70
4.9	Sovrapposizione iniziale con sfere e cuboidi	70
4.10	Analisi dei tempi di esecuzione	71
4.11	Funzione di distribuzione a coppie $g(r)$	73
4.12	Catena di polimeri	74
4.12.1	Catena a incastro libero	74
4.12.2	Flessibilità di una catena di polimeri	75
4.12.3	<i>Persistence Length</i>	75
4.12.4	Stima della <i>persistence length</i> per sfere dure con patch attrattive	77
4.12.5	<i>Persistence length</i> di filamenti di semidischi	79
5	Simulazioni Monte Carlo di Cilindri	83
5.1	Approssimazione a poliedri	83
5.2	Stima della <i>Persistence Length</i> dei cilindri	84
5.3	Energia libera di <i>Stacking</i>	85
5.4	Fasi LC per cilindri con 4 patch	87
5.4.1	Fase nematica	88
5.4.2	Fase colonnare	91
5.4.3	Equazione di stato	93

6	Risultati	95
6.1	Energia libera di <i>staking</i> per semidischi	95
6.2	Stabilità della fase nematica dopo la sostituzione con semidischi . . .	96
6.3	Stabilità della fase colonnare dopo la sostituzione con semidischi . .	99
6.4	Autoassemblamento dei filamenti	104
6.5	Confronto con dati sperimentali	105
7	Conclusioni	111
A	Implementazione e ottimizzazione codice	113
A.1	Patch con esclusività di legame	113
A.2	Linked cell list	114
B	Rilevamento di collisioni - 3D <i>Minkowski Portal Refinement</i>	117
	Bibliografia	121

Introduzione

I cristalli liquidi sono materiali che condividono alcune proprietà strutturali sia dei cristalli che dei liquidi. Possono quindi essere pensati come uno stato intermedio della materia, esistente tra la fase solida e quella liquida, chiamato appunto mesofase. Nei cristalli liquidi, le molecole perdono, completamente o parzialmente, l'ordinamento posizionale presente nei solidi, ma non il loro ordinamento orientazionale (ovvero, come le molecole sono allineate in un campione). Nei liquidi classici, invece, le molecole non hanno un ordinamento posizionale o orientazionale.

I cristalli liquidi furono scoperti per la prima volta nel 1888 dal chimico Friedrich Reinitzer, che osservò che il riscaldamento di un campione di colesteril benzoato (un solido a temperatura ambiente) portava a due punti di fusione. Il primo dei punti di fusione si verificò a $145,5^{\circ}\text{C}$ e portò alla formazione di un liquido torbido. Dopo un ulteriore riscaldamento, si incontrò un secondo punto di fusione a $178,5^{\circ}\text{C}$, al quale il colesteril benzoato si trasformò in un liquido trasparente. Il liquido torbido riportato da Reinitzer è ora noto per essere una fase a cristalli liquidi. Reinitzer inviò una lettera al fisico tedesco Otto Lehmann, insieme a due campioni di colesteril benzoato, in cui chiese un'indagine per confermare le sue osservazioni sperimentali. L'esame di questi campioni ha portato Lehmann a usare l'espressione "cristalli liquidi" per descrivere questi campioni.

Dall'invenzione del primo display a cristalli liquidi nel 1964 presso gli RCA Laboratories di Princeton, nel New Jersey, i cristalli liquidi sono stati al centro di intensa attività di ricerca e altamente interdisciplinare, che comprende chimici, ingegneri, matematici e fisici. Ognuna di queste aree continua a dare un contributo sostanziale allo sviluppo delle tecnologie a cristalli liquidi in un'ampia gamma di applicazioni. Di interesse alla tesi è lo studio delle fasi liquido cristalline di varie tipologie di molecole di DNA. La scoperta che lunghi filamenti di DNA potessero avere fasi liquido cristalline risale agli anni '60, ma solo recentemente, grazie a nuove tecnologie per la sintesi di filamenti di DNA più corti e con forme più complesse, si è ampliata la ricerca per fasi liquido cristalline per varie strutture di DNA. Recentemente la ricerca di fasi liquido cristalline di DNA e RNA ha ricevuto ulteriore interesse dovuto alla teoria per cui queste abbiano avuto un ruolo chiave nell'origine della vita favorendo la formazione di filamenti più lunghi [23].

Si è dimostrato che i trifosfati mononucleotidici dell'acido nucleico (dNTPs e rNTPs), a concentrazione sufficientemente elevata e a bassa temperatura in soluzione acquosa, possono presentare una transizione liquido cristallina a una fase colonnare [53].

Questi filamenti di DNA hanno una lunghezza di circa 4 nm con la possibilità di impilarsi a formare filamenti più lunghi, questa caratteristica permette la formazione della fase colonnare. I risultati sperimentali hanno attirato l'interesse per la costruzione di modelli computazionali che simulino i filamenti di DNA e permettano di prevedere il comportamento del sistema. Troviamo diverse pubblicazioni, in particolare gli articoli [36], [25] e [10], dove si costruiscono modelli *coarse-grained* che vanno a approssimare questi filamenti con cilindri e quasi-cilindri duri con una patch attrattiva su ogni base. In entrambi i sistemi costruiti si trovano delle fasi liquido cristalline nematiche a certe combinazioni di concentrazione e temperatura, queste fasi nematiche però sono assenti negli esperimenti in cui si usano singoli nucleotidi che si autoassemblano.

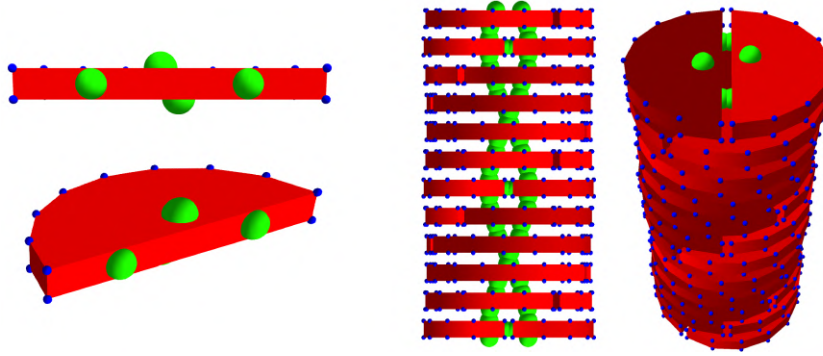


Figura 0.1. (a) Semidisco con 4 patch attrattive, due di *stacking* e due di *Pairing*, in vista laterale. (b) Filamento formato da 24 semidischi con una rotazione di $\pi/5$ tra due vicini.

Da questa discordanza tra fasi liquido cristalline mostrate tra esperimento e simulazione nasce l'idea di questa tesi, si cerca di trovare un modello *coarse-grained* che presenti una fase colonnare ma non una fase nematica. Per andare a considerare anche la flessibilità dei filamenti di DNA si costruisce un modello in cui si approssimano i singoli nucleotidi con dei dischi cilindrici tagliati longitudinalmente a metà e, per tener conto dei legami di *pairing* e di *stacking*, si inseriscono rispettivamente due patch attrattive sul taglio e una patch su ognuna delle due basi circolari come mostrato in figura 0.1. É stato realizzato un programma che effettua una simulazione Monte Carlo nell'*ensemble* NVT o NPT utilizzando l'algoritmo Metropolis [33]. Siccome vengono utilizzate forme geometriche non regolari, un disco tagliato longitudinalmente, è stato necessario trovare un algoritmo per le intersezioni che sia affidabile e veloce visto il gran numero di molecole che andranno simulate. La scelta è ricaduta su un algoritmo che utilizza il separating axes theorem [13] e un altro chiamato *Xenocollide* [16]. I dischi tagliati vengono approssimati a poliedri con un numero di vertici finito, che è possibile scegliere a inizio simulazione, così da poter utilizzare gli algoritmi appena citati. Si procede anche a verificare che l'approssimazione a poliedri non modifichi la fisica delle simulazioni. Oltre a ciò si ottimizzano le simulazioni utilizzando due liste di celle, una per il singolo nucleotide e una per le patch attrattive. Il sistema costruito permette inoltre di modificare la rigidità dei filamenti composti da semidischi, andando a modificare posizione e raggio delle patch attrattive, così da poter adattare la simulazione agli esperimenti con nucleotidi che si autoassemblano.

Il testo della tesi è stato suddiviso nei seguenti capitoli:

- Nel primo capitolo si presenta la teoria dei cristalli liquidi con una introduzione che ne racconta la scoperta e lo sviluppo e in seguito si trova la descrizione delle fasi liquido cristalline di interesse alla tesi.
- Nel secondo capitolo si presenta una parentesi Biofisica sul DNA descrivendo la struttura di quest'ultimo e i possibili cristalli liquidi che può generare in base alla forma e alla lunghezza dei filamenti.
- Nel terzo capitolo si introduce la teoria delle simulazioni Monte Carlo.
- Nel quarto capitolo si descrive l'approccio computazionale alla simulazione, presentando gli algoritmi e le ottimizzazioni utilizzati andando a confrontare i tempi di esecuzione.
- Nel quinto capitolo si analizzano i risultati ottenuti per i cilindri andando a verificare che l'approssimazione a poliedri con un numero di vertici finito non modifichi la fisica del sistema.
- Nel sesto capitolo si presentano i risultati ottenuti andando a simulare il filamento di DNA attraverso i singoli nucleotidi e analizzando la stabilità della fase nematica e della fase colonnare. Si verifica che l'autoassemblamento dei filamenti avvenga correttamente e si confrontano i risultati ottenuti dalle simulazioni con i risultati della Ref. [53] per nucleotidi trifosfato (dNTPs).

Nell'ultima parte conclusiva si riassumono i principali risultati ottenuti proponendo possibili sviluppi futuri. La tesi, quindi, ha come obiettivo quello di realizzare un modello ottimale per simulare al calcolatore filamenti autoassemblanti di DNA.

Capitolo 1

Cristalli liquidi

I cristalli liquidi (LC) sono sistemi anisotropi caratterizzati da un ordinamento parziale, possono essere collocati, in termini termodinamici, tra un solido cristallino tridimensionale e un liquido isotropo [12]. La loro scoperta viene attribuita al botanico austriaco Friedrich Reinitzer che, nel 1888, riferì dell'osservazione di composti (il colesterolo benzoato e il colesterolo acetato) che avevano apparentemente due punti di fusione [46]. Questo capitolo presenta una discussione dettagliata della fasi a cristalli liquidi più comuni: la fase nematica, smettica e colonnare.

1.1 Scoperta e sviluppo

I cristalli liquidi furono scoperti per la prima volta, casualmente, alla fine del diciannovesimo secolo presso la Karl-Ferdinands-Universität dal botanico e chimico austriaco Friedrich Reinitzer (1857–1927). Nel 1888 Reinitzer produceva esteri di colesterolo per degli studi che stava conducendo sulle piante. Studiando le proprietà dei colesteroli estratti dalle carote, notò che il colesterolo benzoato (ovvero il $C_{34}H_{50}O_2$) mostrava non uno, ma due punti di fusione associati a una diversa birifrangenza e a un cambiamento di colore iridescente [46]. Stranamente, il benzoato di colesterolo ottenuto aveva le proprietà ottiche dei cristalli ma scorreva comunque come un liquido.

Reinitzer, non sicuro riguardo la sua scoperta, scrisse una lettera descrivendo l'esperimento al fisico tedesco Otto Lehmann (1855–1922) dell'Università di Aquisgrana. Nella sua lettera, Reinitzer ha affermato che a 145.5 °C, la sostanza sembrava essere un composto torbido e completamente liquido, mentre a 178.5 °C diventava limpido, queste transizioni erano reversibili. Reinitzer chiese a Lehmann, che è stato uno dei primi ricercatori a combinare un microscopio ottico polarizzato con uno a stadio caldo permettendo il controllo della temperatura, che ancora oggi sono considerati apparecchiature standard nella ricerca sui cristalli liquidi, di confermare l'insolito fenomeno dei due punti di fusione. Lehmann, lavorando anche su materiali con un comportamento simile, rispose etichettando le sostanze come "*flüssiger Kristall*" o "cristalli che scorrono" [27]. Pertanto, Lehmann coniò il nome con cui queste sostanze sono ora conosciute: cristalli liquidi.

Reinitzer e Lehmann si resero conto che stavano osservando una nuova fase intermedia della materia, cioè una fase che aveva le proprietà sia di un cristallo che di un liquido. Reinitzer pubblicò quindi quella che è considerata la prima ricerca riconosciuta sui cristalli liquidi: "*Beiträge zur Kenntnis des Cholesterins*" (ovvero "Contributi alla conoscenza del colesterolo") [45]. Successivamente, usando la sua combinazione di microscopio ottico polarizzato e a stadio caldo, Reinitzer riuscì a effettuare osservazioni ad alta temperatura e a controllare le condizioni in cui questi materiali cambiavano fase. Il successivo articolo di Lehmann, "*Über fließende Krystalle*" (ovvero "Su i cristalli liquidi"), presentato nel 1889, aprì un nuovo capitolo della ricerca sui materiali e gettò le basi della scienza dei cristalli liquidi [27].

All'inizio si limitarono i cristalli liquidi a mera curiosità in natura: sostanze strane e peculiari che erano interessanti da osservare ma che non fornivano alcuna applicazione pratica. Tuttavia, la ricerca sui cristalli liquidi è fiorita rapidamente e da allora migliaia di composti organici hanno mostrato questo tipo di comportamento [3]. Nel 1922, il mineralogista e cristallografo francese Georges Friedel (1865–1933) pubblicò "*Les États Mésoformes de la Matière*" (cioè "Stati mesomorfici della materia") in cui proponeva lo schema di classificazione delle fasi cristallo-liquide che è ancora utilizzato oggi [4]. Nel 1927, il fisico russo Vsevolod Fréedericksz (1885–1944) applicò campi esterni per produrre transizioni di fase nei cristalli liquidi. Quando un campo elettrico o magnetico sufficientemente forte viene applicato a un cristallo liquido in uno stato non distorto, si verificherà una transizione di fase quando i componenti del cristallo si allineano al campo, denominata "transizione di Fréedericksz". Le applicazioni, almeno inizialmente, erano molto limitate siccome erano necessarie temperature molto alte per avere le transizioni di fase, lo sviluppo, poi, ha portato le transizioni di fase a temperatura ambiente permettendo quindi lo sviluppo di display con questa tecnologia. La ricerca sui display a cristalli liquidi ha ricevuto un interesse sempre maggiore a partire dagli anni '60. Nella RCA Corporation di Princeton nel 1962, il chimico fisico Richard Williams stava lavorando allo sviluppo di un televisore utilizzando la tecnologia dei cristalli liquidi. L'idea era quella di applicare una tensione a un sottile strato di cristalli liquidi e di generare "zone" o "domini" di motivi a strisce mentre il campione cambia il suo allineamento in risposta al campo applicato. Questi "domini di Williams" erano i precursori dei pixel che si trovano nei display a cristalli liquidi odierni. Successivamente il primo schermo LCD utilizzabile a temperatura ambiente fu realizzato grazie alla scoperta, ad opera di George William Gray, della fase nematica del 4-ciano-4'-pentilbifenile ($C_{18}H_{19}N$), liquido cristallino trasparente con anisotropia dielettrica positiva che presenta una fase nematica stabile a temperatura ambiente. Questo composto viene tutt'oggi utilizzato nei display a tecnologia *twisted nematic LCD* a colori grigi e neri come in calcolatrici e orologi digitali.

Negli anni '90, la ricerca sui cristalli liquidi si era evoluta in un campo ricco con un suo potenziale per far progredire lo sviluppo scientifico e le sue applicazioni innovative oltre all'ambito delle tecnologie di visualizzazione. A testimonianza di quanti sviluppi ci siano stati nel breve tempo dalla scoperta di Reinitzer, il fisico francese Pierre-Gilles de Gennes (1932–2007) ha ricevuto il Premio Nobel per la Fisica nel 1991 per il suo lavoro proprio sui cristalli liquidi [9].

1.1.1 Nomenclatura dei cristalli liquidi

La nomenclatura dei cristalli liquidi è soggetta a modifiche più o meno continue in base ai sviluppi nella ricerca scientifica. Inizialmente si aveva una classificazione delle fasi in nematica, colesterica e smettica, introdotta da Friedel [17] nel 1922. Si classificavano le fasi riferendosi a sistemi con solo ordine orientazionale a lungo raggio (nemática), la fase nemática elicoidale osservata per composti derivati dalla colesterina (colesterico) e sistemi con ordine posizionale addizionale, quindi fasi con struttura a strati (smettico). Le fasi venivano contraddistinte per le trame tipiche ottenute nella microscopia polarizzata. Vale anche la pena notare a questo punto che le fasi nemática e colesterica sono sostanzialmente equivalenti dal punto di vista termodinamico. Quest'ultima è semplicemente una fase nemática composta da molecole chirali, che mostra una sovrastruttura elicoidale.

Successivamente, a seguito della scoperta di fasi intermedie, fu introdotta la nomenclatura che assegnava delle lettere a fasi diverse. Il loro schema non seguiva una classificazione sistematica come nel caso delle strutture cristalline, ma le fasi erano fondamentalmente nominate in ordine dalla loro scoperta. Pur offrendo uno schema di classificazione inequivocabile, questa procedura aveva creato una certa confusione all'inizio, poiché gruppi diversi nominavano fasi diverse con la stessa lettera, o le stesse fasi con lettere diverse, che avrebbero dovuto essere assegnate a un tipo di fase già noto. Molti di questi problemi vennero risolti quando Sackmann e Demus proposero la loro regola di miscibilità [48] per l'identificazione delle fasi a cristalli liquidi: se una fase sconosciuta è continuamente miscibile con una fase di riferimento sull'intero intervallo di concentrazione, allora sono equivalenti. D'altra parte, se non sono miscibili, non si può trarre alcuna conclusione. Ciò implica che deve essere determinato il diagramma di fase completo tra l'incognita e il composto di riferimento.

Si utilizza nella tesi la nomenclatura di Lagerwall [26] ovvero si userà l'asterisco per denotare una fase composta da molecole chirali, indipendentemente da ogni possibile sovrastruttura elicoidale. Ciò significa che non si userà la notazione che si trova spesso "Ch" per colesterico, ma piuttosto N^* .

1.2 Fasi liquido cristalline

La proprietà più interessante dei cristalli liquidi è la varietà di mesofasi che spesso esibiscono. Queste diverse mesofasi sono classificate in base alla quantità di ordine che possiedono. Un sistema sarà un cristallo liquido se presenterà fasi intermedie della materia come una fase nemática o una fase smettica. Pertanto, una grande varietà dei materiali è stata classificata come cristalli liquidi, ad esempio l'acido desossiribonucleico (DNA), come discusso nel prossimo capitolo ma anche vari polimeri. Come possono componenti così diverse essere in grado di presentarsi con fasi della materia comprese tra quella di un solido e quella di un liquido? Nonostante le differenze significative nella composizione chimica e nelle proprietà di questi composti, questi materiali condividono una caratteristica comune che si presta alla formazione di fasi cristalline liquide: una forma altamente anisotropa delle molecole o degli aggregati. Infatti, una forma estesa, strutturalmente rigida, anisotropa dei

componenti di base è la radice di quasi tutti i sistemi liquidi cristallini [20]. In questo capitolo si discute come vengono osservate sperimentalmente queste mesofasi e si presentano le fasi principali dei cristalli liquidi.

1.2.1 Birifrangenza e analisi sperimentale dei cristalli liquidi

Per osservare e caratterizzare le varie mesofasi che si possano formare, si utilizza il microscopio ottico a luce polarizzata, apparecchiatura utilizzata anche inizialmente da Reinitzer ma che ancora oggi è in uso nella versione *hot-stage* che permette il controllo della temperatura del campione. Come discusso in precedenza le mesofasi sono caratterizzate da un cambiamento nella birifrangenza, quest'ultima non è altro che la dipendenza dalla direzione dell'interazione della luce con la sostanza. In altre parole, se l'interazione della luce con un materiale dipende dalla direzione, o anche polarizzazione, della luce incidente, quel materiale è detto birifrangente [20]. Quando la luce entra in un mezzo, la sua velocità e lunghezza d'onda vengono ridotte di un fattore adimensionale n , chiamato indice di rifrazione. L'indice di rifrazione è definito come il rapporto tra la velocità della luce nel vuoto e la velocità della luce nel mezzo considerato. Ad esempio, l'indice di rifrazione del cloruro di sodio NaCl, è circa pari a $n \approx 1.54$, indicando che la luce viaggia nel sale a una velocità di:

$$c_{NaCl} \approx \frac{c_{vuoto}}{1.54} = \frac{2.9979 \times 10^8 \text{ ms}^{-1}}{1.54} = 1.947 \times 10^8 \text{ ms}^{-1} \quad (1.1)$$

Il cloruro di sodio è una sostanza isotropa quindi non ha assi di simmetria unici nel reticolo cristallino quindi la luce interagisce allo stesso modo indipendentemente dall'orientazione del cristallo rispetto al fascio di luce incidente. Al contrario i cristalli liquidi sono sostanze anisotrope, questa caratteristica si manifesta sperimentalmente attraverso la birifrangenza del materiale. I cristalli liquidi più semplici sono uniassiali, il che significa che l'anisotropia esiste lungo un'unica direzione, chiamata asse ottico, mentre tutte le direzioni perpendicolari all'asse ottico sono equivalenti tra loro.

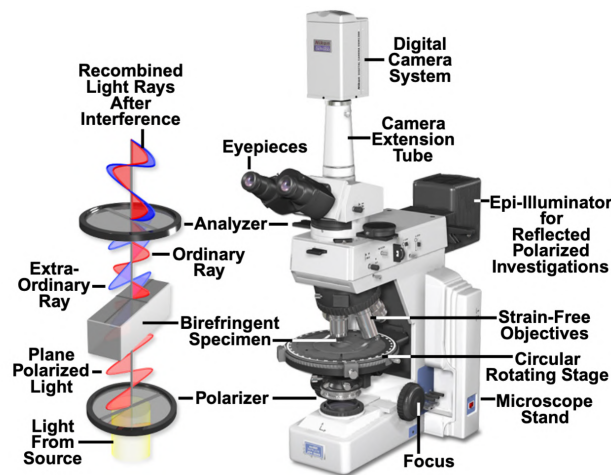


Figura 1.1. Schema di funzionamento di un microscopio ottico polarizzato. Immagine tratta da: microscopyu.com/techniques/polarized-light

La birifrangenza dei cristalli liquidi si presenta come cambiamenti di colore e di intensità di luce quando vengono analizzati utilizzando un microscopio ottico polarizzato, rappresentato in figura (1.1). Il fascio di luce passa prima in un polarizzatore per poi raggiungere il campione, che essendo birifrangente, andrà a rompere il fascio in due singole componenti d'onda, ognuna polarizzata su differenti piani e ognuna con una propria velocità di propagazione. Passato il campione, i fasci di luce passano attraverso un secondo polarizzatore. Per un materiale posto tra due polarizzatori l'intensità della luce trasmessa sarà proporzionale a $(\sin(\phi/2))^2$ dove $\phi = 2\pi d(\Delta n)/\lambda$ è lo sfasamento dovuto al campione, Δn è l'indice di birifrangenza del materiale e d è lo spessore del materiale. Se si considera un materiale isotropo, ovvero con birifrangenza nulla $\Delta n = 0$, l'intensità sarà sempre nulla e quindi il campione apparirà completamente nero se analizzato con un microscopio ottico polarizzato.

Successivamente nello studio si ruota il campione per modificare l'orientazione tra asse ottico e raggio di luce incidente e si modificano le fasi del materiale controllando la temperatura del materiale. Quando il campione liquido cristallino viene ruotato, l'orientamento relativo dell'asse ottico e del raggio di luce incidente cambia. Nel frattempo, mentre il campione passa attraverso varie fasi (a causa delle variazioni di temperatura nel caso di un sistema termotropico), il grado di ordine associato a ciascuna fase cambia. Entrambi i fenomeni modificano le proprietà ottiche della porzione di campione attraverso la quale passa la luce.

1.2.2 Fase solida e fase liquida

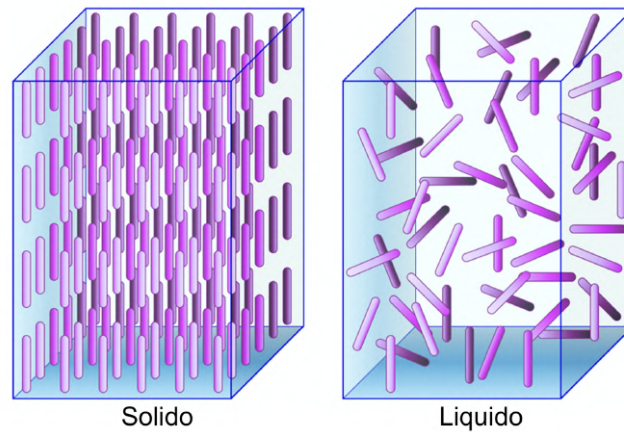


Figura 1.2. Nella fase solida, nella figura a sinistra, le molecole sono disposte su un reticolo tridimensionale. Nella fase liquida, nella figura a destra, le molecole non presentano nessuna periodicità spaziale e sono libere di muoversi. Immagine tratta da [20].

Dato un materiale, le differenze tra lo stato solido e quello liquido, mostrate in figura (1.2), derivano dal grado di ordine con cui sono disposte le molecole che lo compongono. Ovviamente lo stato solido è caratterizzato da un alto grado di ordine di posizione e di orientazione. I centri di massa delle molecole, o di gruppi di molecole, occupano posizioni su un reticolo tridimensionale. Mentre nella fase

liquida generalmente i centri di massa degli elementi costitutivi non hanno alcuna periodicità tridimensionale.

1.2.3 Fase isotropa

La fase isotropa, raggiunta a alte temperature, è caratterizzata come una fase liquida infatti in questo caso il materiale è libero di scorrere e le molecole non presentano ordine sia nelle posizioni che nelle orientazioni. Come già discusso il sistema avrà birifrangenza nulla e risulterà nero sotto analisi con luce polarizzata.

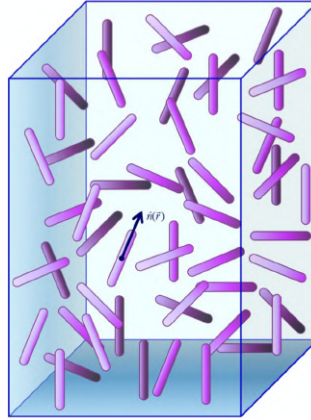


Figura 1.3. Rappresentazione schematica della fase isotropa dove le molecole sono disposte in posizioni e orientazioni casuali. Immagine tratta da [20].

In figura (1.3) si mostra uno schema che rappresenta la configurazione isotropa, con $\hat{n}(\vec{r})$ che indica la direzione di elongazione della molecola con centro di massa in \vec{r} . Nel caso isotropo non ci sono direzioni favorite e di conseguenza la media spaziale di $\hat{n}(\vec{r})$ sarà nulla. Il sistema non ha un ordine a lungo raggio e la funzione di correlazione $G(\vec{r}) = \langle S(0)S(\vec{r}) \rangle$ è proporzionale a $e^{-r/\xi}$, dove ξ è la lunghezza di correlazione isotropa mentre S è il parametro ordine nematico scalare che verrà discusso meglio nel prossimo paragrafo. Al diminuire della temperatura del sistema le molecole iniziano a presentare dell'ordine sulle lunghe distanze così da far passare a una nuova fase il sistema.

1.3 Fase nematica

Nella fase nematica le molecole, in media, sono orientate lungo una particolare orientazione \hat{n} , chiamata asse nematico, questa descrive l'allineamento nematico del campione. Mentre nelle orientazioni c'è un ordinamento, per quanto riguarda i centri di massa delle molecole queste non hanno un ordine sulle lunghe distanze, di conseguenza i centri di massa si comportano come un liquido convenzionale. La direzione \hat{n} è scelta casualmente nello spazio, nella maggior parte dei casi è imposta dalla forma del volume in cui avviene la simulazione, ma può essere selezionata imponendo un campo magnetico esterno nella prima parte della simulazione. Sperimentalmente sono disponibili varie tecniche per scegliere la direzione di allineamento.

Nella fase nematica ogni molecola deve essere identica alla sua immagine speculare, o, se è presente una molecola chirale (cioè una molecola destrorsa o sinistrorsa diversa dalla sua immagine speculare), il campione deve essere racemico (cioè contenente una miscela 1-a-1 di molecole destre e sinistre). La fase colesterica, che verrà discussa brevemente, descrive sistemi le cui molecole sono chirali e sono mescolate in un rapporto non racemico. Uno schema della fase nematica è mostrata in figura (1.4).

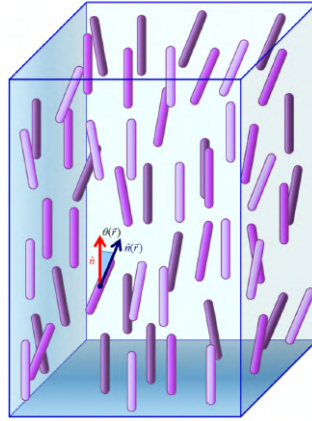


Figura 1.4. Una rappresentazione schematica della fase nematica dove le molecole sono disposte in posizioni casuali ma presentano una orientazione media comune. Immagine tratta da [20].

1.3.1 Parametro d'ordine nematico

Nell'analisi dei cristalli liquidi è necessario descrivere l'ordine orientazionale senza l'ordine posizionale, quindi bisogna costruire un parametro di ordine orientazionale per la fase nematica. Per fare ciò è necessario costruire un parametro di ordine orientazionale per la fase nematica a cristalli liquidi. Per costruire un parametro di ordine orientazionale, si utilizza una media sulle orientazioni di tutte le molecole in un insieme. Quindi, si rappresenta l'orientamento di ciascuna molecola come un vettore unitario \hat{n} . Questo vettore unitario ha componenti n_α , con $\alpha = x, y$ e z [52]. Come primo tentativo, si può costruire un parametro di ordine orientazionale facendo la media di tutti i vettori unitari nell'insieme, ovvero:

$$\mathbf{M} = \langle \hat{n} \rangle \quad (1.2)$$

Nella fase isotropa \mathbf{M} è in media zero. Nella fase nematica, poiché le molecole hanno la stessa probabilità di puntare verso l'alto o verso il basso lungo l'asse principale, \mathbf{M} ha una media nulla, proprio come nella fase isotropa. Per questo motivo, il parametro d'ordine \mathbf{M} non può descrivere l'ordinamento orientazionale nella fase nematica. È necessario trovare una quantità che tenga conto solo della direzione e non del verso della molecola. Si potrebbe utilizzare il prodotto tensoriale medio, si definisce ovvero il parametro d'ordine tensoriale:

$$\mathbf{T} = \langle \hat{n} \otimes \hat{n} \rangle \equiv \langle \hat{n} \hat{n} \rangle \quad (1.3)$$

Il primo controllo è verificare che questo tensore riesca a distinguere la fase isotropa e quella nematica. Per farlo si valuta prima il tensore nella fase isotropa. Le componenti diagonali sono $T_{xx} = \langle n_x^2 \rangle$, $T_{yy} = \langle n_y^2 \rangle$ e $T_{zz} = \langle n_z^2 \rangle$. Queste tre componenti devono essere uguali tra loro perché la fase è isotropa, inoltre devono sommarsi a 1 perché \hat{n} è un vettore unitario. Quindi, ogni componente diagonale deve essere $1/3$. Le componenti fuori diagonale sono $T_{xy} = T_{yx} = \langle n_x n_y \rangle$, $T_{xz} = T_{zx} = \langle n_x n_z \rangle$ e $T_{zy} = T_{yz} = \langle n_z n_y \rangle$. Poiché la fase è isotropa, i tre componenti n_x , n_y e n_z hanno la stessa probabilità di essere positivi o negativi e fluttuano indipendentemente l'uno dall'altro. Quindi, tutti i componenti fuori diagonale mediano a 0. Mettendo insieme questi risultati, il tensore diventa nella fase isotropica:

$$\mathbf{T}_{\alpha\beta} = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix} = \frac{1}{3} \delta_{\alpha\beta} \quad (1.4)$$

Successivamente si valuta il tensore in uno stato con perfetto ordine nematico lungo l'asse z . In questo stato, tutte le molecole hanno l'orientamento $\hat{n} = \pm \hat{z}$. Di conseguenza, la componente $T_{zz} = 1$ e tutte le altre componenti sono nulle. Quindi, il tensore diventa:

$$\mathbf{T}_{\alpha\beta} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (1.5)$$

Quindi il tensore $\mathbf{T}_{\alpha\beta}$ può distinguere tra la fase isotropa e quella nematica. Tuttavia, si apportano due piccole modifiche al parametro dell'ordine. In primo luogo, l'Eq. 1.4 mostra che la media isotropa di $\mathbf{T}_{\alpha\beta}$ è diversa da zero. È più conveniente avere un parametro d'ordine nullo nella fase isotropa. Possiamo ottenerlo sottraendo la media isotropa, per ottenere $\mathbf{T}'_{\alpha\beta} = \mathbf{T}_{\alpha\beta} - \frac{1}{3} \delta_{\alpha\beta}$. Quindi, $\mathbf{T}'_{\alpha\beta} = 0$ nella fase isotropa, mentre nella fase nematica ordinata lungo l'asse z :

$$\mathbf{T}'_{\alpha\beta} = \begin{bmatrix} -\frac{1}{3} & 0 & 0 \\ 0 & -\frac{1}{3} & 0 \\ 0 & 0 & \frac{2}{3} \end{bmatrix} \quad (1.6)$$

Per evitare di avere la fase nematica con il termine $\frac{2}{3}$, si moltiplica tutto il tensore per un fattore $\frac{3}{2}$ ovvero $\mathbf{Q}_{\alpha\beta} = \frac{3}{2} \mathbf{T}'_{\alpha\beta}$. Ricapitolando il tensore ottenuto è:

$$\mathbf{Q}_{\alpha\beta} = \left\langle \frac{3}{2} n_\alpha n_\beta - \frac{1}{2} \delta_{\alpha\beta} \right\rangle \quad (1.7)$$

Per costruzione nella fase isotropa $\mathbf{Q}_{\alpha\beta} = 0$, mentre nella fase nematica allineata sull'asse z sarà pari a:

$$\mathbf{Q}_{\alpha\beta} = \begin{bmatrix} -\frac{1}{2} & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (1.8)$$

Il tensore ha due proprietà matematiche rilevanti, è simmetrico, il che significa che $Q_{\alpha\beta} = Q_{\beta\alpha}$, e è a traccia nulla $Tr(\mathbf{Q}) = Q_{\alpha\alpha}$.

Si consideri ora un sistema con un ordinamento nematico solo parziale lungo l'asse z . Ogni molecola avrà il proprio vettore unitario $\hat{\mathbf{n}}$ che è possibile scrivere in coordinate sferiche come $\hat{\mathbf{n}} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$. Dove θ è l'angolo rispetto l'asse z e varia tra 0 e π ; ϕ è l'angolo azimutale nel piano xy e sarà distribuito uniformemente tra 0 e 2π . Di conseguenza le medie azimutali sono $\langle \cos \phi \rangle = \langle \sin \phi \rangle = \langle \cos 2\phi \rangle = \langle \sin 2\phi \rangle = 0$ e $\langle \cos^2 \phi \rangle = \langle \sin^2 \phi \rangle = 1/2$. Il tensore di ordine nematico sarà:

$$Q_{zz} = \left\langle \frac{3}{2} \cos^2 \theta - \frac{1}{2} \right\rangle = \langle P_2(\cos \theta) \rangle \quad (1.9)$$

dove $P_2(u) = \frac{3}{2}u^2 - \frac{1}{2}$ è il polinomio di secondo grado di Legendre mentre $S = Q_{zz}$ è il parametro d'ordine scalare della fase nematica. Le altre componenti del tensore d'ordine nematico sono:

$$\begin{aligned} Q_{xx} &= \left\langle \frac{3}{2} \sin^2 \theta \cos^2 \phi - \frac{1}{2} \right\rangle = \left\langle \frac{3}{4} \sin^2 \theta - \frac{1}{2} \right\rangle = -\frac{S}{2} \\ Q_{yy} &= \left\langle \frac{3}{2} \sin^2 \theta \sin^2 \phi - \frac{1}{2} \right\rangle = \left\langle \frac{3}{4} \sin^2 \theta - \frac{1}{2} \right\rangle = -\frac{S}{2} \\ Q_{xz} &= Q_{zx} = \left\langle \frac{3}{2} \cos \theta \sin \theta \cos \phi \right\rangle = 0 \\ Q_{yz} &= Q_{zy} = \left\langle \frac{3}{2} \cos \theta \sin \theta \sin \phi \right\rangle = 0 \\ Q_{xy} &= Q_{yx} = \left\langle \frac{3}{2} \sin^2 \theta \cos \phi \sin \phi \right\rangle = \left\langle \frac{3}{4} \sin^2 \theta \sin 2\phi \right\rangle = 0 \end{aligned} \quad (1.10)$$

Di conseguenza il tensore che rappresenta il parziale ordine nematico nella direzione dell'asse z :

$$Q_{\alpha\beta} = \begin{bmatrix} -\frac{S}{2} & 0 & 0 \\ 0 & -\frac{S}{2} & 0 \\ 0 & 0 & S \end{bmatrix} \quad (1.11)$$

Nel caso l'ordinamento fosse in direzione x o y allora il termine S sarà nella rispettiva posizione. Quando l'asse nematico non è parallelo a uno degli assi del sistema di riferimento è necessario modificare il calcolo del tensore; si riscriva l'equazione 1.11 come:

$$Q_{\alpha\beta} = S \left[\frac{3}{2} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right] = S \left[\frac{3}{2} z_\alpha z_\beta - \frac{1}{2} \delta_{\alpha\beta} \right] \quad (1.12)$$

dove $z_\alpha = (0, 0, 1)$, quindi se l'orientazione è verso una generica direzione $\hat{\mathbf{l}}$:

$$Q_{\alpha\beta} = S \left[\frac{3}{2} l_\alpha l_\beta - \frac{1}{2} \delta_{\alpha\beta} \right] \quad (1.13)$$

come in precedenza $S = \langle P_2(\cos \theta) \rangle$ e θ è l'angolo che forma da $\hat{\mathbf{l}}$.

Il primo fattore S dell'equazione 1.13, chiamato parametro dell'ordine scalare, descrive il livello dell'ordine nematico. Mostra quanto bene le molecole sono allineate

tra loro, infatti se $S = 0$, non c'è allineamento e il sistema è in una fase isotropa. Se $S = 1$, c'è allineamento completo e il sistema è in uno stato nematico perfetto. In generale, per i tipici cristalli liquidi nematici, S è compreso tra 0 e 1 e il sistema ha un ordine nematico parziale [52].

Il secondo fattore $[\frac{3}{2}l_\alpha l_\beta - \frac{1}{2}\delta_{\alpha\beta}]$ descrive la direzione dell'ordine nematico. Il vettore \hat{l} , con componenti n_α , è detto asse nematico. È un vettore unitario che rappresenta l'asse principale dell'ordine nematico. Riassumendo, per caratterizzare una fase nematica, è necessario calcolare l'autovettore associato all'autovalore più grande del tensore d'ordine nematico e questi saranno rispettivamente l'asse nematico e il parametro d'ordine nematico.

1.3.2 Fase colesterica o nematica chirale N^*

La fase colesterica è molto simile alla fase nematica, infatti, la fase colesterica è spesso indicata come fase nematica* o fase N^* . La differenza principale è che i componenti molecolari sono chirali nella fase colesterica. Analogamente alla fase nematica, le molecole costituenti nella fase colesterica sono libere di scorrere l'una accanto all'altra in tutte le direzioni e mancano di ordine posizionale nei loro centri di massa. Tuttavia, a differenza della normale fase nematica il sistema perde un piano speculare di simmetria [7]. Una molecola è chirale se la sua immagine speculare non può essere sovrapposta alla sua immagine originale.

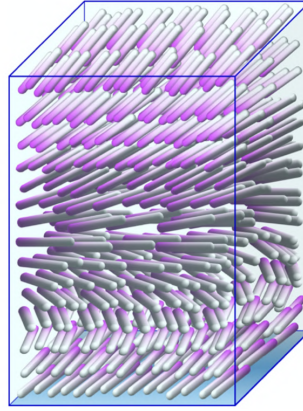


Figura 1.5. Una rappresentazione schematica della fase nematica chirale dove le molecole sono state disegnate con punte chiare e scure in modo da poter visualizzare la torsione del direttore. Immagine tratta da [20].

Nella fase colesterica, la perdita della simmetria speculare fa sì che il direttore molecolare \hat{n} , non sia una costante nello spazio, ma la conformazione preferita è quella elicoidale. Il direttore molecolare ora diventa una funzione di z (ovvero $\hat{n} \rightarrow \hat{n}(z)$), quindi le molecole cambiano orientazione spostandosi sull'asse z come mostrato nella figura (1.5). Pertanto, \hat{n} cambia in funzione di z secondo:

$$\hat{n}(z) = (n_x(z), n_y(z), 0) = (\cos(q_0 \cdot z + \phi_0), \sin(q_0 \cdot z + \phi_0), 0) \quad (1.14)$$

Dove $q_0 = \pi/d$ è il vettore d'onda e ϕ_0 è la fase. Il passo di questa torsione elicoidale è definito come la distanza necessaria al direttore per compiere una rotazione completa ed è determinato dalla composizione chimica della molecola e dalla temperatura. Per $q_0 > 0$, il passo dell'elica è destrorso. Per $q_0 < 0$, il passo dell'elica è sinistrorso. Inoltre, il passo dell'elica è altamente sensibile alla temperatura: all'aumentare della temperatura il passo diminuisce.

1.4 Fase Smettica

Diminuendo la temperatura, il campione di cristalli liquidi nematici entra in una delle numerose fasi stratificate, le più comuni delle quali sono le fasi smettica A e smettica C. Si noti che le fasi smettiche, che sono più ordinate della fase nematica, si verificano sempre a temperature inferiori all'intervallo di temperatura nematica. Si discutono quindi separatamente le fasi smettiche A, B e C.

1.4.1 Fase Smettica A

Nella fase smettica A è stato introdotto un grado di ordine traslazionale che era assente nella fase nematica. Le molecole sono ancora orientate preferenzialmente lungo la direzione indicata da \hat{n} , tuttavia, le molecole sono ora disposte in strati la cui normale è parallela a \hat{n} . Per la maggior parte, le molecole sono limitate al movimento all'interno di uno strato.

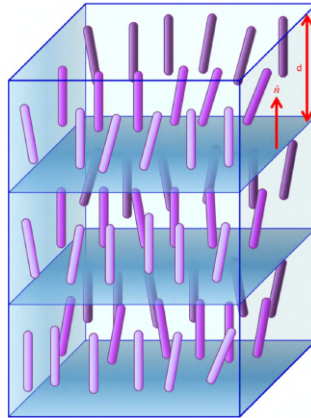


Figura 1.6. Rappresentazione schematica della fase Smettica A dove le molecole sono allineate a strati con l'asse lungo delle molecole diretto perpendicolarmente ai piani degli strati. Immagine tratta da [20].

La struttura a strati della fase smettica A ha tipicamente altezza tra layer pari alla lunghezza della molecola costituente, nel caso di dimeri si hanno strati spessi due lunghezze molecolari. All'interno di ogni strato, le molecole non mostrano un ordine a lungo raggio. Ogni strato si comporta come un liquido bidimensionale. Il sistema è otticamente uniaassiale con l'asse ottico che è la normale agli strati (ovvero lungo \hat{n}). Le direzioni $+\hat{z}$ e $-\hat{z}$ sono totalmente equivalenti, con \hat{z} si indica la normale allo strato. Uno schema della fase smettica A è mostrato in figura 1.6.

1.4.2 Fase Smettica B

Come accennato in precedenza, nella fase smettica A , le molecole sono orientate perpendicolarmente allo strato smettico, ma nello strato non esiste un ordine a lungo raggio. In sostanza, ogni strato si comporta come un liquido bidimensionale. D'altra parte, nella fase smettica B , è presente un ordine all'interno degli strati smettici. Le molecole sono disposte in una rete di schemi esagonali all'interno degli strati, che non sono allineati tra loro, e quindi rendono la fase smettica B la più ordinata delle fasi smettiche. La fase esattico smettica B è illustrata nella figura 1.7.

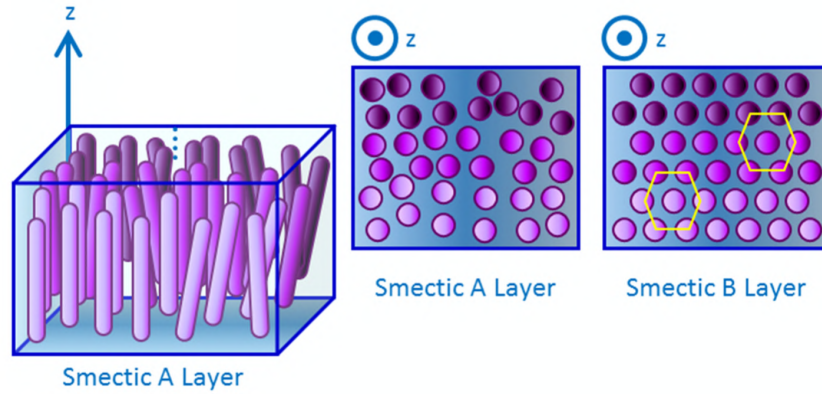


Figura 1.7. Rappresentazione schematica che mostra la differenza tra la fase smettica A e smettica B , nella seconda, guardando il sistema sull'asse \hat{z} , è possibile notare un ordinamento esagonale. Immagine tratta da [20].

1.4.3 Fase Smettica C

Nel momento in cui la temperatura viene abbassata ulteriormente nella fase smettica A può avvenire una transizione di fase in cui viene mantenuta la struttura stratificata ma le molecole subiscono un'inclinazione θ_T rispetto la normale al layer.

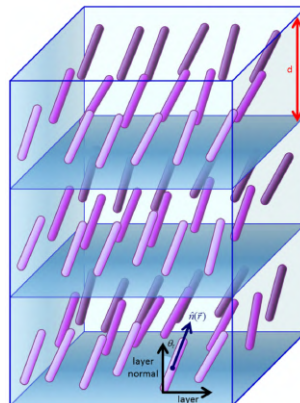


Figura 1.8. Rappresentazione schematica della fase Smettica C dove le molecole sono allineate a strati con l'asse lungo delle molecole inclinato di θ_T rispetto ai piani degli strati. Immagine tratta da [20].

Questa fase è denominata smettica *C*. Come nel caso della smettica *A*, in ogni strato, le molecole non presentano un ordinamento a lungo raggio. A differenza della fase smettica *A* la smettica *C* è otticamente biassiale. La struttura smettica *C* si ottiene solo quando le molecole costituenti sono otticamente inattive. Se, infatti, nella fase smettica *C* sono presenti molecole otticamente attive, la struttura si distorcerà in maniera simile a quella che si verifica nella fase nematica. Uno schema della fase smettica *C* è mostrato in figura 1.8.

1.5 Fase Colonnare *Col*

La fase colonnare ha la caratteristica di rompere l'invarianza traslazionale lungo due direzioni. Il sistema è composto da una serie di colonne posizionate su un piano cristallino bidimensionale, che può essere esagonale o cubico, sul piano perpendicolare all'asse direttore. Questa struttura è mostrata schematicamente in figura 1.9. Lungo l'asse \hat{z} le particelle si comportano come un fluido, permettendo alle colonne di scorrere tra di loro a causa della mancanza di un ordine traslazionale in questa direzione.

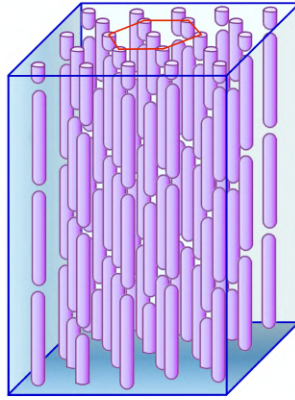


Figura 1.9. Rappresentazione schematica della fase colonnare dove le molecole si allineano lungo l'asse direttore formando colonne, che rimangono libere di scorrere tra loro. Le colonne si dispongono su un piano cristallino esagonale.

1.5.1 Parametro d'ordine esagonale

Per distinguere una fase colonnare da una fase nematica è utile introdurre un nuovo parametro che quantifichi l'ordinamento esagonale sul piano ortogonale all'asse nematico. Viene quindi definito il parametro d'ordine esagonale $\langle \psi_6 \rangle$:

$$\langle \psi_6 \rangle = \left\langle \frac{1}{N} \sum_{i=1}^N \frac{1}{n(i)} \sum_{j=1}^{n(i)} e^{6i\theta_{ij}} \right\rangle \quad (1.15)$$

Dove $n(i)$ è il numero dei primi vicini alla molecola i sullo stesso strato, θ_{ij} è l'angolo tra le due molecole i e j rispetto a un asse ortogonale all'asse nematico \hat{n} come mostrato in figura 1.10. La costante presente a esponente definisce la periodicità valutata dal parametro, in questo caso $k = 6$ nel caso esagonale, ma

è possibile utilizzare anche $k = 4$ per quantificare un ordinamento quadratico. Il parametro d'ordine esagonale restituisce un valore non nullo in caso della presenza di un ordine esagonale e è nullo nel caso opposto.

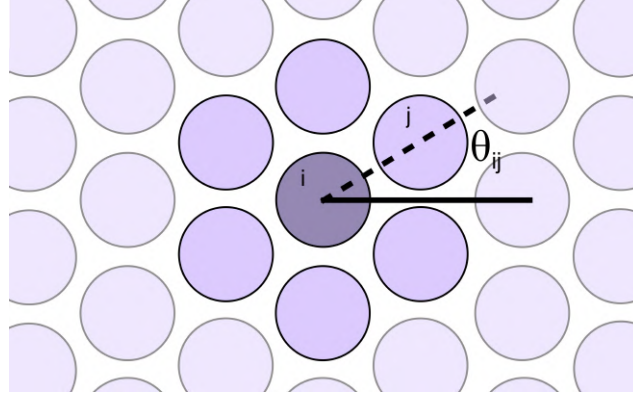


Figura 1.10. Vista superiore dell'ordinamento esagonale. La linea tratteggiata unisce i primi vicini i e j e forma l'angolo θ_{ij} con l'asse \hat{x} . La proiezione delle molecole è stata fatta su un piano perpendicolare all'asse nematico \hat{n} .

Capitolo 2

Parentesi Biofisica: DNA

2.1 Struttura chimica degli acidi nucleici

Gli acidi nucleici, DNA e RNA, sono costituiti da catene polinucleotidiche, cioè polimeri lineari di unità chiamate nucleotidi. Si descrivono prima i singoli nucleotidi isolati per poi descrivere come questi andranno a legarsi in una catena di acido nucleico. I nucleotidi sono molecole composte da tre componenti: una base azotata, uno zucchero pentoso e uno o più gruppi fosfato [15].

Nel DNA lo zucchero pentoso presente è il deossiribosio, mentre nell'RNA è il ribosio. I due zuccheri, come mostrato in figura (2.1), differiscono per un gruppo OH presente sulla posizione 2 del ribosio e assente nel deossiribosio. Questa differenza è molto rilevante, in quanto questo gruppo OH conferisce instabilità alle molecole di RNA, infatti il DNA, che ha un H in quella stessa posizione, è metabolicamente più stabile.

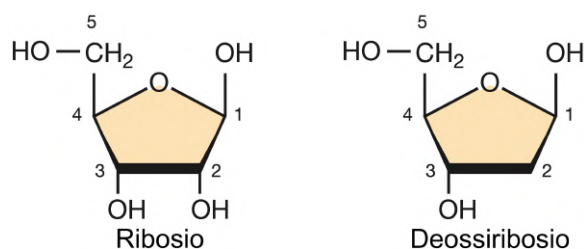


Figura 2.1. Il ribosio è lo zucchero a cinque atomi di carbonio che si trova nell'RNA. Il deossiribosio si trova nel DNA e ha un ossigeno in meno rispetto al ribosio poiché ha idrogeno al posto del gruppo idrossile nella posizione 2' dell'anello del ribosio [15].

Gli acidi nucleici naturali possono essere formati da due tipi di basi azotate: purine (a doppio anello eterociclico) e pirimidine (a singolo anello). Nella figura (2.2) si mostrano le specifiche basi che si trovano negli acidi nucleici naturali: due purine, adenina e guanina (A e G), presenti sia nel DNA che nell'RNA, e tre tipi di pirimidine, citosina, timina e uracile (C, T e U). La citosina è comune a DNA e RNA, mentre la timina è componente del solo DNA trovandosi al suo posto l'uracile nell'RNA. La differenza tra uracile, presente nel DNA, e timina, presente invece nell'RNA, è la presenza di un gruppo metilico CH₃ in posizione 5 dell'anello pirimidinico [8].

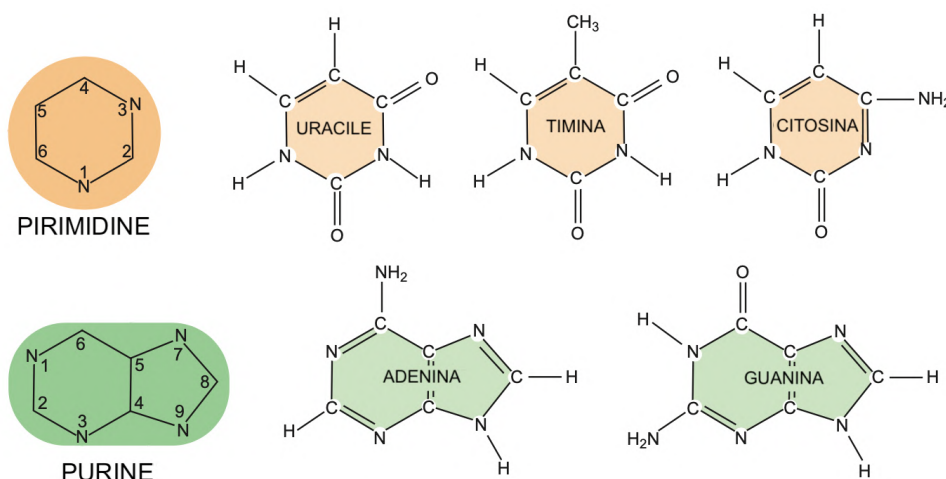


Figura 2.2. Le quattro basi del DNA sono adenina, guanina, citosina e timina. Nell'RNA, l'uracile sostituisce la timina. Le basi pirimidiniche contengono strutture a un anello, mentre le basi puriniche contengono strutture a due anelli.

Quando una delle suddette basi è legata nella posizione 1 di uno zucchero (ribosio o deossiribosio), abbiamo i nucleosidi, che prendono il nome di adenosina, guanosina, citidina, timidina e uridina. Ai nucleosidi possono essere legati uno o più gruppi fosfato PO_4^{3-} e, in tal caso, prendono il nome di nucleotidi: questi sono chiamati acido adenilico, acido guanilico e via dicendo. Il gruppo fosfato conferisce una valenza acida alla molecola e può essere legato al carbonio 5 dello zucchero oppure al carbonio 3, come mostrato in figura (2.3).

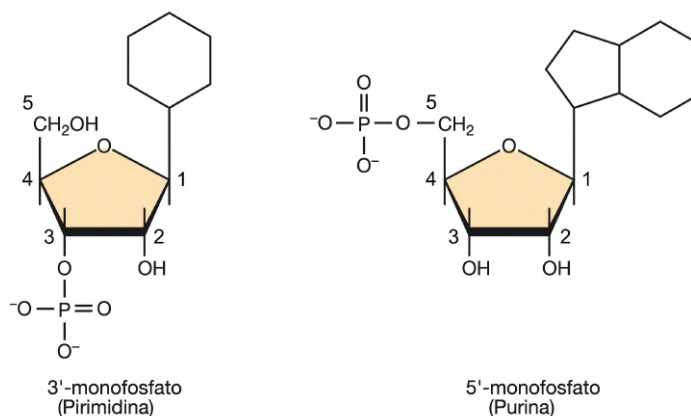


Figura 2.3. I nucleotidi possono avere il gruppo fosfato legato in posizione 5' o 3'. Immagine tratta da [15].

Inoltre, considerati come molecole indipendenti (cioè non inserite nella catena polinucleotidica), i nucleotidi possono contenere uno solo, due o tre gruppi fosfato. Per riassumere la nomenclatura, in figura 2.4 si può vedere che una base legata allo zucchero si definisce nucleoside e se il nucleoside è legato al fosfato si chiama nucleotide.

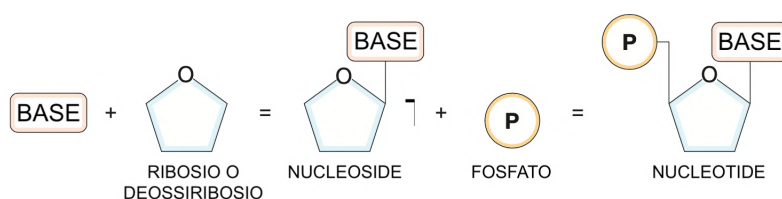


Figura 2.4. Una base legata allo zucchero costituisce un nucleoside. Se il nucleoside è legato a uno o più gruppi fosfato viene detto nucleotide.

Si discute ora come i nucleotidi andranno a legarsi tra loro per formare una catena polinucleotidica di DNA o di RNA. In figura (2.5) si mostra che lo scheletro della catena polinucleotidica è costituito dall'alternanza di zuccheri (deossiribosio o ribosio) e di gruppi fosfato, e che le basi sporgono lateralmente da questo scheletro. Ogni base è legata alla posizione 1' di uno zucchero da un legame glicosidico.

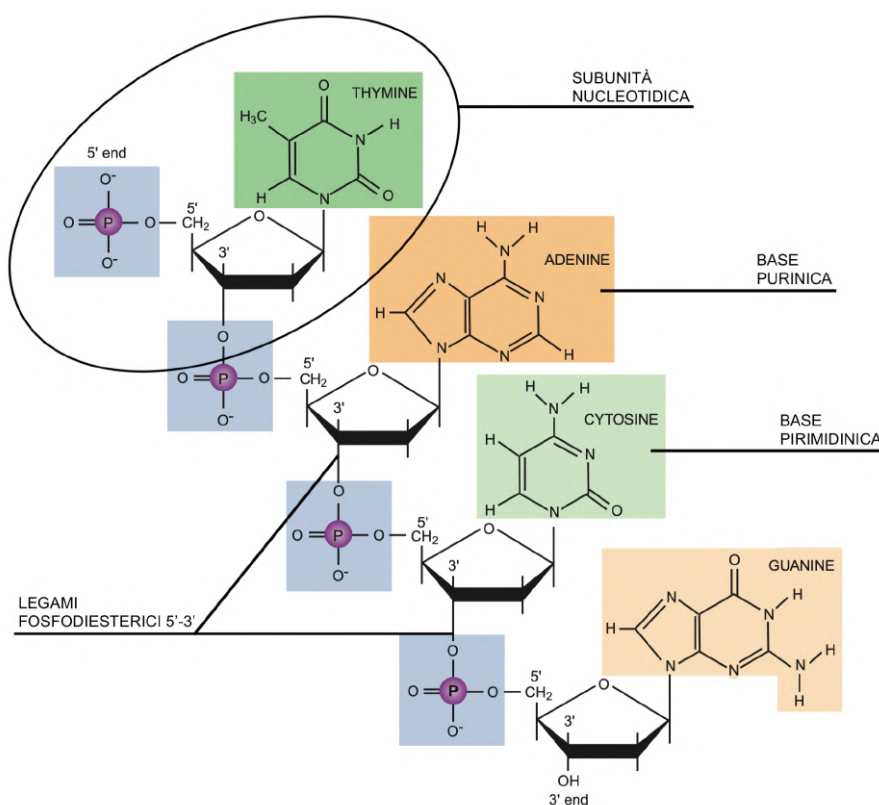


Figura 2.5. Catena polinucleotidica con struttura ripetitiva che consiste in uno scheletro in cui si alternano zuccheri e gruppi fosfato, legati tra loro da legami 5' → 3' fosfodiesterici. Da questa impalcatura protrudono lateralmente le basi azotate. Immagine tratta da [8].

Come evidenziato nella figura, ciascun gruppo fosfato forma un legame estere con il carbonio 5' di uno zucchero e un secondo legame estere con il carbonio 3' dello zucchero successivo. Questo tipo di legame, chiamato fosfodiesterico, conferisce una polarità alla catena polinucleotidica, ovvero questa presenta due diverse estremità: da una parte c'è un nucleotide che ha il carbonio 5' libero, mentre il carbonio

3' è impegnato nel legame fosfodiesterico con il nucleotide adiacente; dall'altra la molecola polimerica termina con un nucleotide che ha il carbonio 5' impegnato nel legame fosfodiesterico con quello adiacente, mentre il carbonio 3' è libero. I carbonio 5' e 3' liberi vuol dire che i C⁻ terminali non sono impegnati in legami fosfodiesterici; la loro valenza è comunque impegnata da gruppi OH⁻ o da gruppi fosfato.

La polarità 5' → 3' delle catene polinucleotidiche, sia di DNA che di RNA, è molto importante perché la struttura dello scheletro della catena polinucleotidica è diversa se vista da un estremo o dall'altro e questo permette tutte le reazioni di sintesi e copiatura, e anche la maturazione e la degradazione degli acidi nucleici.

2.2 Struttura fisica del DNA

Per comprendere la funzione di una macromolecola non è sufficiente conoscerne la struttura chimica, ma è necessario conoscerne la struttura tridimensionale. La tecnica che si è dimostrata fondamentale per risolvere la struttura del DNA è stata la diffrazione dei raggi X, che veniva già utilizzata sugli studi sulle proteine. La tecnica della diffrazione dei raggi X richiede che la macromolecola che si vuole analizzare possa essere ottenuta in forma di cristallo in cui le molecole, e quindi gli atomi che le compongono, sono tutte allineate in maniera ordinata.

Mentre era relativamente facile ottenere piccoli cristalli di proteine purificate, questo non si rivelò possibile per il DNA. Tuttavia, anche se non si tratta di un vero cristallo, il DNA può essere ridotto in forma di fibre in cui le lunghe molecole sono orientate circa parallelamente l'una all'altra. Queste fibre di molecole orientate permisero di ottenere delle immagini di diffrazione dei raggi X che, anche se non ottimali, consentirono di ricavare alcuni parametri strutturali.

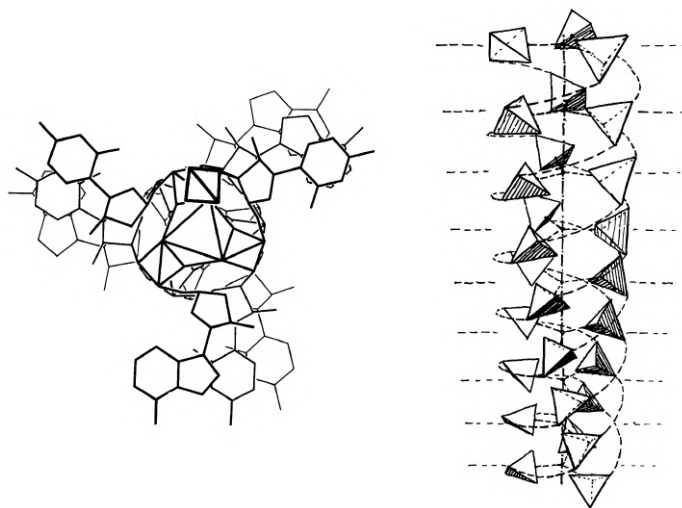


Figura 2.6. Struttura errata a tre eliche proposta inizialmente da Pauling [42], a sinistra la vista dall'alto e a destra la vista laterale.

Una delle prime strutture, che in seguito si rivelò errata, fu proposta da Pauling nel 1952 [42], questa consisteva in un DNA composto da tre eliche intrecciate tra loro con lo scheletro zucchero-fosfato all'interno e le basi sporgenti verso l'esterno. Lo schema originale di Pauling è mostrato in figura (2.6). Questo particolare delle basi rivolte all'esterno era stato proposto per giustificare il fatto che i dati sperimentali indicavano che gli scheletri zucchero-fosfato delle eliche formano una struttura regolare con diametro abbastanza costante, cosa difficile da spiegare se le basi, pirimidine più piccole e purine più grandi, si fossero trovate all'interno della struttura.

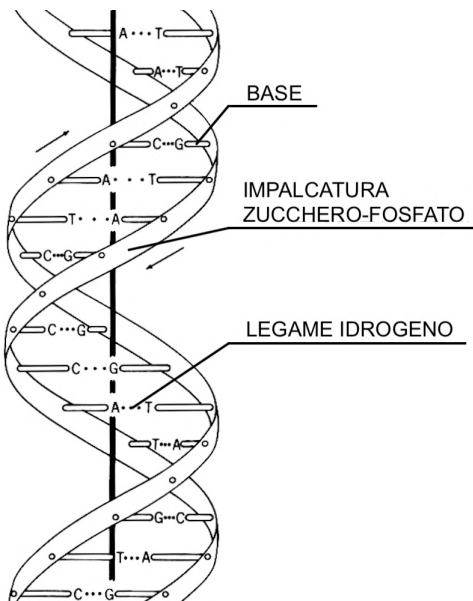


Figura 2.7. Struttura DNA teorizzata da Francis Crick e James Watson [58].

Successivamente, grazie alle immagini di diffrazione a raggi X prodotte dalla ricercatrice Rosalind Franklin, Francis Crick e James Watson capirono che il DNA dovesse essere formato da un'elica. Inoltre, erano convinti che la struttura di Pauling con i fosfati al centro dell'elica fosse sbagliata e quindi decisero di disporre i due filamenti zucchero-fosfato all'esterno e di sistemare le basi affacciate all'interno della struttura. Questa soluzione, insieme all'idea di mettere di fronte una A a una T e una C a una G, permetteva allo stesso tempo di giustificare la regolarità del diametro della doppia elica (ogni coppia di basi è costituita da una purina e una pirimidina) e di dare una spiegazione logica alla regola di Chargaff (la quantità di A è uguale alla quantità di T e la quantità di G è uguale alla quantità di C). Riuscirono così a trovare la struttura fisica del DNA.

La figura (2.7) mostra la figura originale dell'articolo che Watson e Crick pubblicarono sulla rivista scientifica *Nature* il 25 aprile 1953 [58].

2.3 Parametri strutturali del DNA

Le coppie di basi, considerati i loro anelli eterociclici, hanno una struttura praticamente piatta e sono disposte quasi perpendicolarmente all'asse della doppia elica come mostrato in figura (2.8). Le coppie di basi che collegano i due filamenti sono sempre una pirimidina e una purina, in particolare si trova sempre l'appaiamento tra una A e una T e tra una G e una C, questo giustifica la regolarità del diametro della doppia elica.

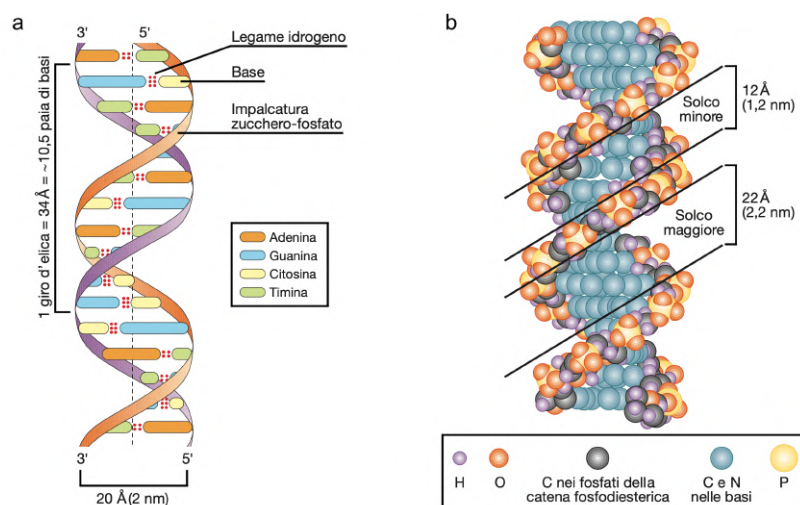


Figura 2.8. Rappresentazione grafica della doppia elica di DNA, in (a) una rappresentazione schematica e in (b) la struttura molecolare a atomi pieni. Immagine tratta da [15]

Le basi si legano tra loro attraverso i legami idrogeno, in particolare nella coppia di basi A-T la A si trova di fronte alla T con una geometria e una distanza tali da permettere la formazione di due legami idrogeno, invece nella coppia G-C si formano tre legami idrogeno. La geometria che permette la formazione dei legami idrogeno è mostrato in figura (2.9).

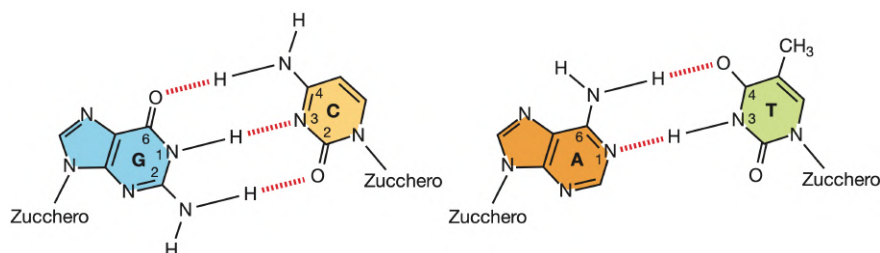


Figura 2.9. Legami idrogeno tra le coppie di basi del DNA, nell'immagine superiore la G e la C formano tre legami idrogeno e nell'immagine inferiore la A e la T ne formano due. Immagine tratta da [15]

La doppia elica del DNA compie un giro completo ogni 34 \AA e ha un diametro di circa 20 \AA . La distanza tra due basi vicine è pari a 3.4 \AA , di conseguenza ci sono circa 10 coppie di basi per ogni giro dell'elica. Come mostrato in figura (2.8) c'è una asimmetria che genera due solchi di dimensione diversa, appunto uno minore e uno maggiore, questa è una conseguenza della posizione delle molecole di deossiribosio ai lati delle basi. A causa dell'opposta polarità $5' \rightarrow 3'$ delle due eliche, i due zuccheri di ciascuna coppia di nucleotidi si vengono a trovare dallo stesso lato dell'elica, come mostrato in figura 2.9. Questa disparità nei due solchi è importante per il riconoscimento della sequenza di DNA: il solco maggiore ha molti più atomi che possono essere donatori o accettori di legami idrogeno e come conseguenza il "linguaggio" presentato da questa zona dell'elica è molto più vario e ricco che nel solco minore.

2.4 Stabilità della doppia elica di DNA in soluzione

In vista delle discussioni che seguiranno nella tesi si discute qui la stabilità della molecola di DNA in soluzione acquosa. La catena singola polinucleotidica ha una porta idrofilica, ovvero lo zucchero-fosfato, e una parte idrofobica, le basi. Nella doppia elica i due scheletri zucchero-fosfato, idrofilici, con le cariche negative dei fosfati, si dispongono all'esterno della struttura a contatto con l'acqua, mentre le basi, che sono idrofobiche, vanno a disporsi all'interno non andando quindi a contatto con l'acqua. La stabilità della doppia elica, cioè la propensione a mantenere le due catene polinucleotidiche legate tra loro, dipende da varie forze. I due scheletri che contengono gli atomi di fosforo carichi negativamente tendono a respingersi; questa repulsione viene controbilanciata da due tipi di forze che tendono a mantenerle appaiate. Innanzitutto i legami idrogeno tra le basi, che sono tre tra ogni coppia G-C e due tra ogni coppia A-T. La stabilità è aumentata dall'effetto idrofobico dovuto all'impilamento (stacking) delle basi.

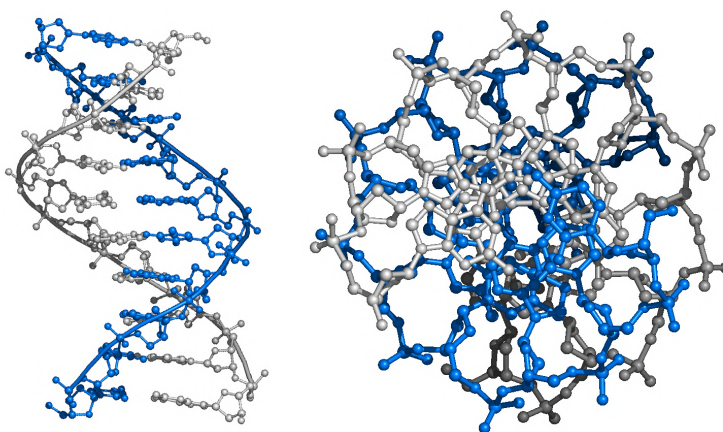


Figura 2.10. Vista laterale e vista dell'alto di un filamento di DNA parziale.

Guardando lungo l'asse dell'elica, come mostrato in figura (2.10), si vede come le coppie di basi adiacenti presentino una considerevole sovrapposizione, anche se non completa a causa della rotazione dell'elica. Questa sovrapposizione è stabilizzata dalla idrofobicità delle basi stesse, che, così impilandosi, mantengono minimo il contatto con l'acqua. Si sottolinea inoltre che c'è una differenza di energia se sopra una coppia di AT c'è una coppia di GC o di CG, quindi la geometria di impilamento ha un'influenza da un punto di vista energetico.

Da quanto detto deriva che la stabilità della doppia elica in soluzione non è uguale per tutti i DNA. Inoltre, dipende dalla particolare sequenza di nucleotidi nel DNA: infatti, l'effetto idrofobico dipende dall'impilamento e quindi dalla sovrapposizione di coppie di basi adiacenti e quanto due coppie di basi sono sovrapposte dipende dalla particolare sequenza di nucleotidi. Proprio a causa delle interazioni di impilamento

tra le basi e ai gradi di flessibilità che hanno i legami chimici dei nucleotidi, il parallelismo tra due coppie di nucleotidi lungo l'asse dell'elica non è quasi mai perfetto, come si può vedere in figura (2.10), e quindi la relativa posizione delle basi è stata catalogata in vari parametri.

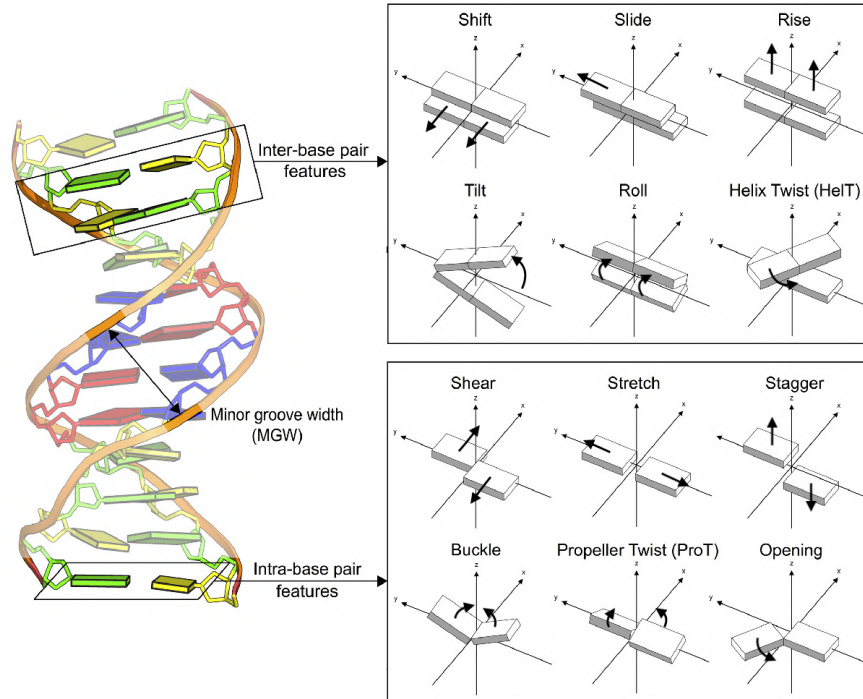


Figura 2.11. Schematizzazione dei dodici parametri che caratterizzano la posizione reciproca delle basi. L'asse \hat{z} corrisponde all'asse della doppia elica di DNA e l'asse \hat{y} la direzione dei legami a idrogeno. Immagine tratta da [28].

Come mostrato in figura 2.11 i parametri che descrivono la posizione reciproca delle basi sono in totale dodici, divisi in due gruppi da sei, le interazioni tra due coppie di basi e le interazioni interne a due basi accoppiate. Il parametro di Twist rappresenta l'angolo tra due coppie di nucleotidi rispetto all'asse della doppia elica, mediamente di 34° che causa la rotazione dell'elica di DNA ogni 10.4 basi. Il Roll è l'inclinazione, dovuta alle forze di repulsione tra due coppie di nucleotidi, rispetto all'asse della coppia AT o GC e il Tilt è l'angolo che si forma, sempre per repulsione, rispetto all'asse dello scheletro zucchero-fosfato. Queste posizioni relative, e quindi i diversi possibili valori dei dodici parametri, influenzano la forma stessa del DNA, che può diventare "curvo" in base alla ripetizione di particolari sequenze nucleotidiche.

2.5 Forze di Pairing e Stacking

La sequenza delle basi azotate è ciò che conferisce al DNA la sua capacità di trasportare l'informazione genetica, per questo è fondamentale che i nucleotidi formino catene di legami covalenti [2]. Al contrario, i due filamenti sono tenuti insieme da interazioni deboli per consentire la chiusura e l'*unfolding* della doppia elica, il passaggio critico coinvolto nella "lettura" della sequenza durante la trascrizione e la

replicazione del DNA nella cellula. In effetti, il ruolo biologico del DNA è radicato nelle proprietà fisiche della sua struttura molecolare. La solubilità, la flessibilità e le interazioni reciproche dei filamenti di DNA sono fondamentali per la sua capacità di codificare e trasferire informazioni biologiche. La forma stessa della doppia elica è il risultato di un sottile equilibrio di vincoli molecolari e forze di interazione, che giocano tutti un ruolo nei processi biologici. Come già detto le doppie eliche sono tenute insieme da una combinazione di due principali forze intermolecolari: le cosiddette forze di accoppiamento di basi (Pairing) e di impilamento di basi (Stacking).

Queste forze controllano la selettività del processo di legame, per cui l'energia di legame dipende dal grado di complementarità delle sequenze di nucleobasi lungo i due polimeri. Infatti, se si effettua un'analisi dell'energia libera, si nota che il legame è quasi trascurabile nell'accoppiamento C-G, mentre nel caso delle coppie A-T l'energia è addirittura positiva, quindi il legame è sfavorito. Il risultato è mostrato in figura 2.12 (a), tratta da Ref. [60], si nota inoltre che non c'è dipendenza dalla temperatura e quindi che non è presente un contributo entropico significativo. Comunque le coppie non-WC, ovvero tutte le coppie che non seguono le regole di appaiamento di Watson e Crick, risultano più sfavorite.

In figura 2.12 (b) viene mostrata l'energia libera di Stacking in funzione della temperatura, come discusso nel paragrafo precedente, questa risulta essere il motivo principale per cui le catene polinucleotidiche rimangano unite e è dovuta per la maggior parte alle basi che tendono a sovrapporsi per minimizzare il contatto con l'acqua essendo idrofobiche.

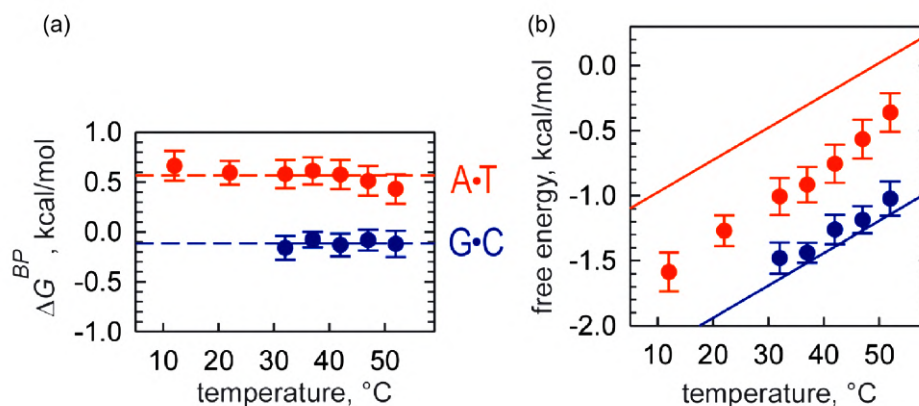


Figura 2.12. Contributi di energia libera di Pairing e Stacking di basi alla stabilità termica dell'elica. In blu il contributo dovuto alla coppia G-C e in rosso quello dovuto alla coppia A-T. (a) il grafico dell'energia libera dovuta al legame idrogeno non mostra dipendenza dalla temperatura. (b) Lo stacking favorisce la formazione del duplex con una netta dipendenza dalla temperatura. Immagine tratta da [60].

Un filamento di DNA porta una carica negativa per ciascun gruppo fosfato in un'ampia gamma di valori di pH e la doppia elica nuda ha quindi un'elevata densità di carica lineare di circa $6 e^-/nm$, quindi andare a modificare la soluzione in cui si trova il DNA, cambiando la forza ionica, comporta una modifica nell'energia dei suoi

legami. Sebbene nelle soluzioni elettrolitiche lo screening dei controioni possa ridurre significativamente la densità di carica effettiva [56], la repulsione elettrostatica gioca un ruolo significativo nella stabilità dell'elica del DNA.

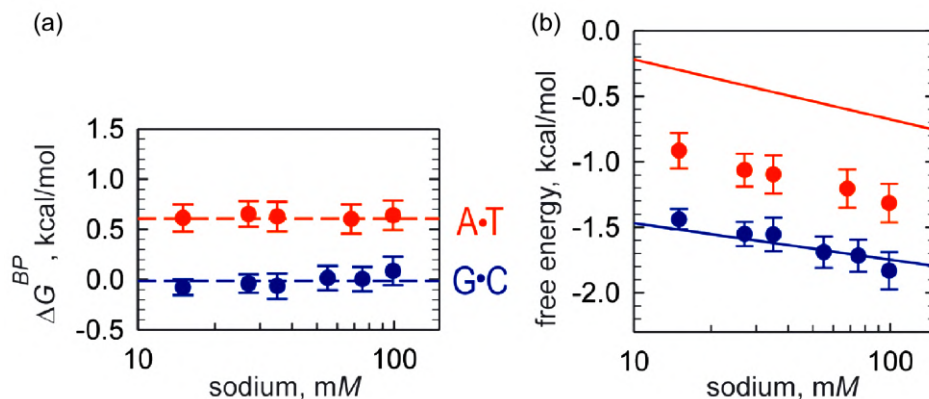


Figura 2.13. Dipendenza dell'energia libera di ibridazione del DNA dalla forza ionica. L'impilamento di basi (a) è la parte principale dell'interazione mentre l'accoppiamento di basi (b) è un contributo piccolo dipendente dalla densità del sale. Immagine tratta da [60].

Nell'articolo di John SantaLucia e Donald Hicks [50] e in lavori precedenti, si presume che la dipendenza dal sale provenga solo dal contributo entropico, mentre l'entalpia sarebbe indipendente dal sale [49]. Tuttavia, misurazioni successive hanno riportato che sia ΔH che ΔS aumentano con l'aumento della forza ionica e quindi entrambi contribuiscono all'aumento di energia libera osservato. Con lo stesso sistema sperimentale sopra descritto, si è studiato come l'energia libera di stacking e di paring dipendono dalla quantità di sale in soluzione [60]. In figura 2.13 si mostra che la forza ionica non influenza l'accoppiamento delle basi mentre influenza l'impilamento.

2.6 Cristalli liquidi di DNA

La modularità e la selettività delle interazioni DNA-DNA forniscono uno strumento straordinariamente flessibile per la progettazione di aggregati con struttura controllata e interazioni reciproche, grazie alla possibilità di sintetizzare chimicamente oligonucleotidi di determinate sequenze. La sintesi del DNA, sviluppata negli anni '80 [6], viene effettuata su substrati in fase solida utilizzando la fosforamidite, la procedura viene schematizzata in figura 2.14. Questa tecnica permette la costruzione di oligomeri di lunghezza fino a 100 paia di basi in qualche milligrammo di quantità. Si avrà un processo di interazione a due stadi, inizialmente si sfrutta l'accoppiamento WC per produrre gli aggregati fortemente legati, nei casi più semplici si costruisce la classica doppia elica di DNA altrimenti si possono costruire aggregati multifilari, che saranno localmente delle doppie eliche, ma che globalmente potranno formare strutture più complesse. Gli aggregati formati andranno poi a interagire tra loro in diversi modi.

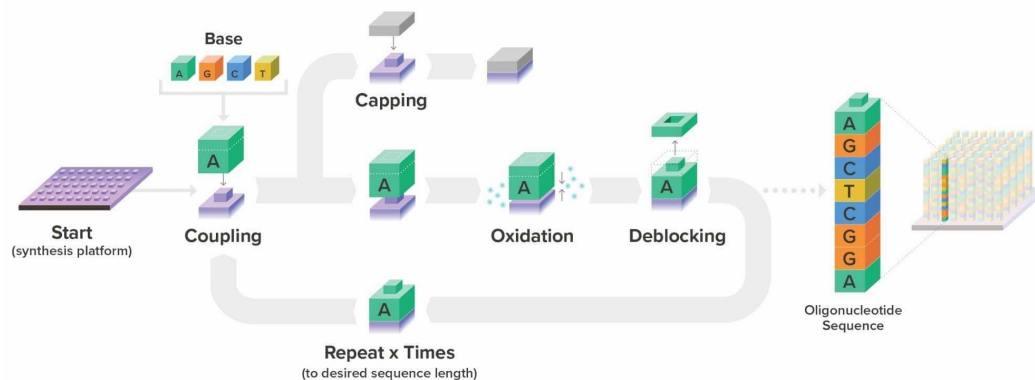


Figura 2.14. Schema della procedura per la sintesi di oligomeri, questa tecnica permette di decidere la sequenza delle basi. Il ciclo di *Coupling*, *Oxidation* e *Deblocking* si ripete fino a costruzione dei filamenti e in conclusione si utilizza la tecnica del *Capping* per evitare che le sequenze disaccoppiate potrebbero creare errori nella molecola sintetizzata. Immagine tratta da www.twistbioscience.com.

I cristalli liquidi (LC) sono fasi fluide caratterizzate da un parziale ordinamento orientazionale e posizionale tra lo stato liquido isotropo e la struttura cristallina. Anche se agli inizi venivano visti come una strana forma di ordinamento collettivo di un insieme molto limitato di molecole organiche, i LC sono oggi riconosciuti come una forma comune di autorganizzazione della materia soffice e biologica. Le fasi LC del DNA sono state osservate per la prima volta in vitro con lunghi filamenti doppi e recentemente hanno iniziato a essere considerati un ordinamento comune per gli oligomeri del DNA.

2.6.1 Filamenti di DNA lunghi

I primi studi condotti tra il 1950 e il 1960 identificarono il comportamento liquido cristallino di lunghe catene di DNA in vitro [47]. Poco dopo Bouligand iniziò un'analisi sistematica delle textures di vari campioni biologici ottenuti mediante microscopia ottica ed elettronica, in particolare dei cromosomi di Dinoflagellates (alghe unicellulari primitive) trovando evidenza di fasi liquido cristalline di DNA altamente concentrato in vivo [5]. Le fasi del DNA LC sono state ampiamente studiate utilizzando sistemi sperimentali più facilmente controllabili, vale a dire doppi filamenti di DNA ottenuti attraverso il taglio enzimatico del DNA nucleosomico, fornendo un insieme di $N \sim 150$ bastoncini semirigidi, o attraverso la sonicazione.

Sono state identificate due mesofasi principali:

1. Una fase **nematica chirale (N*)**, chiamata anche fase colesterica. Tale fase liquido cristallina è costituita da "strati" nematici locali in cui l'asse nematico ruota con continuità, con periodicità $p/2$ (dove p è il passo colesterico come mostrato in figura 2.15a [30]. Tale fase LC è facilmente osservabile in microscopia ottica polarizzata, poiché la *pitch* si estende a decine di micrometri e l'anisotropia ottica delle basi del DNA porta a strutture caratteristiche come le goccioline in figura 2.15b.

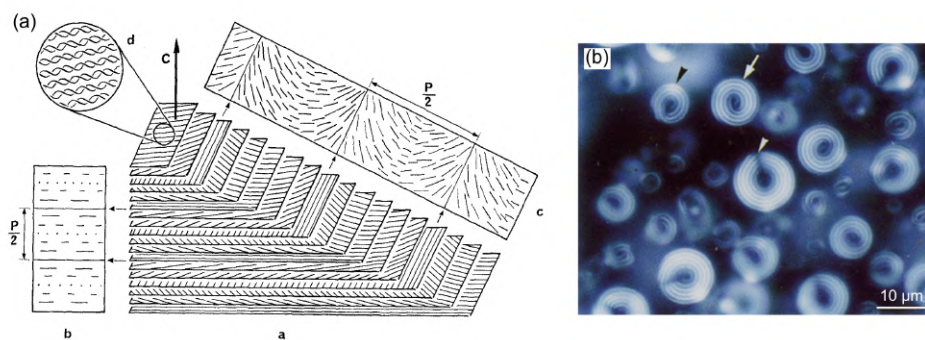


Figura 2.15. (a) Rappresentazione schematica della struttura della fase nematica chirale del DNA. (b) N^* goccioline osservate nella microscopia polarizzata. Le strisce chiare e scure corrispondono a $p/2$. Immagini tratte da [30].

2. Una **fase colonnare** (COL), in cui eliche di DNA parallele si allineano su un reticolo 2D (come mostrato in figura 2.16a) ma rimangono disordinate nella direzione ortogonale. La continua flessione delle colonne dà origine ai cosiddetti domini sviluppabili, mostrati in 2.16b. La natura della fase e la sua simmetria di impaccamento esagonale sono state dimostrate da una combinazione di esperimenti di microscopia polarizzata, diffrazione di raggi X e microscopia elettronica a frattura da congelamento ¹ [62]. La fase COL si osserva per concentrazioni superiori a 400 mg/mL.

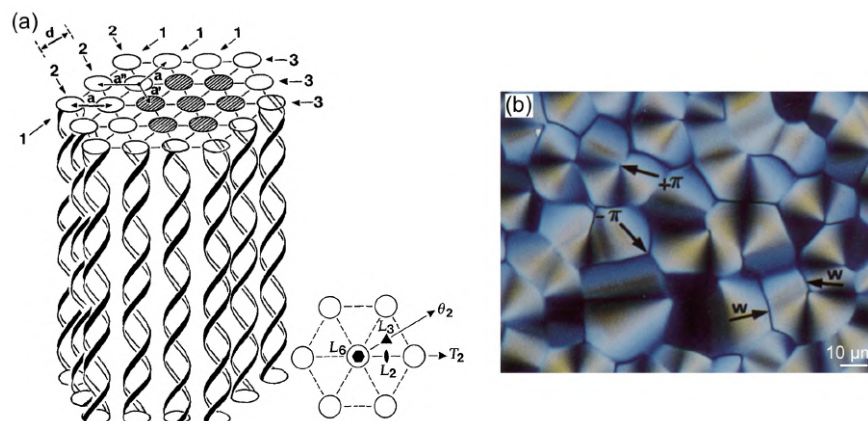


Figura 2.16. (a) Rappresentazione della struttura della fase colonnare esagonale del DNA, che mostra molecole parallele impaccettate in modo esagonale. (b) Domini sviluppabili COL osservati nella microscopia polarizzata. "w" indica le pareti del difetto tra domini orientati in modo diverso, mentre " $\pm\pi$ " sta per difetto puntuale attorno al quale le molecole di DNA si piegano continuamente. Immagini tratte da [30].

¹La tecnica del congelamento-frattura consiste nel rompere fisicamente (fratturare) un campione biologico congelato. Il dettaglio strutturale esposto dal piano di frattura viene quindi visualizzato mediante microscopio elettronico a trasmissione. La frattura da congelamento è l'unica tra le tecniche di microscopia elettronica nel fornire viste planari dell'organizzazione interna delle membrane.

L'aspetto di una fase nematica del DNA deriva dalla sua forma astiforme e dalla rigidità quando è in conformazione elicoidale. Sebbene il DNA a filamento singolo sia estremamente flessibile, con una lunghezza di persistenza del polimero di appena 2 nm [55], quando accoppiato nella doppia elica diventa un oggetto piuttosto rigido, con una lunghezza di persistenza di circa 50 nm (circa 150 paia di basi) [31].

Secondo l'argomentazione classica di Onsager [39] basata sulle interazioni di volume escluso, ovvero lo spazio negato alle molecole dovuto alla presenza di altre molecole, ci si aspetta che i bastoncini repulsivi semiflessibili a una concentrazione sufficientemente alta sviluppino un ordine orientazionale. L'ordine nematico diventa favorito perché la perdita di entropia orientazionale è più che compensata dall'aumento dell'entropia posizionale. Nonostante la sua polidispersità, flessibilità e repulsione elettrostatica, il DNA si è dimostrato un buon banco di prova per la previsione di Onsager [39]: una volta che il diametro effettivo D viene adeguatamente ridimensionato per includere interazioni repulsive (a loro volta dipendenti dalla forza ionica), la frazione di volume critica per l'ordinamento nematico corrisponde alla previsione di Onsager, $\phi_{IN} \approx 4D/L$ [32], dove L è la lunghezza della doppia elica. Il comportamento, in termini di fasi, per filamenti di diversa lunghezza è mostrato in figura 2.17.

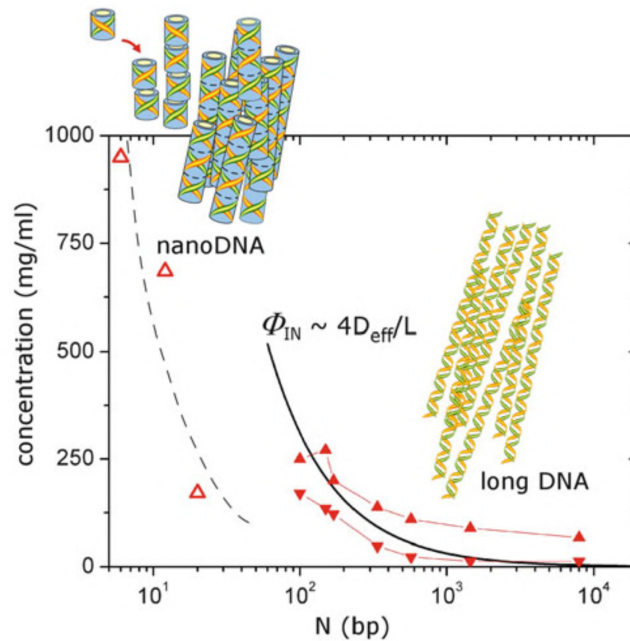


Figura 2.17. Comportamento di fase di molecole di DNA di diversa lunghezza. Le eliche più lunghe di 100 bp (triangoli pieni) mostrano una transizione da isotropica a nematica a concentrazioni ragionevolmente ben descritte dalla teoria di Onsager una volta che il diametro viene ridimensionato per la repulsione elettrostatica (linea nera continua). Al contrario, i filamenti di DNA corti mancano dell'anisotropia necessaria per allinearsi spontaneamente e le loro concentrazioni critiche (triangoli aperti) implicano l'impilamento in aggregati più lunghi. I dati del DNA lungo e del nanoDNA provengono rispettivamente da [32] e [35]. Il grafico è tratto dalla Ref. [2].

In un sistema hard-rod, a una frazione di volume sufficientemente elevata è solitamente prevista una transizione dalla fase nematica alla fase smettica A [4], una fase lamellare con strati perpendicolari al direttore nematico. Tuttavia, come dimostrato da Livolant [30], nel DNA la fase smettica è sostituita dall'ordinamento colonnare; questo comportamento può essere facilmente spiegato sulla base della flessibilità del filamento [52] o della polidispersità della lunghezza [3], entrambi che favoriscono la fase COL rispetto a quella smettica. L'alta concentrazione, tuttavia, non è l'unico modo per ottenere fasi LC del DNA in soluzioni acquose.

È stato riscontrato che il DNA tende a aggregarsi e a precipitare dopo l'aggiunta nelle soluzioni di vari agenti condensanti, che introducono interazioni efficaci interelicoideali attrattive. Questo accade con alcoli e altri solventi, che riducono la solubilità del DNA [43], o con cationi multivalenti come spermidina, spermina e cobalto esammina.

2.6.2 Filamenti di DNA corti

Secondo la teoria di Onsager e le simulazioni al computer del comportamento di sferocilindri duri [4], in assenza di interazioni aggiuntive non è previsto alcun ordinamento LC per *aspect-ratio* $L/D < 4$, e quindi doppie eliche di DNA con un numero di coppie di basi $N < 24$ mancherebbero dell'anisotropia per mostrare il comportamento liquido cristallino a qualsiasi concentrazione (figura 2.17). Tuttavia l'ordinamento LC è stato recentemente trovato in soluzioni acquose concentrate di sequenze autocomplementari di DNA con N a partire da 6 bp ("nano-DNA", nDNA), che si accoppiano in eliche con proporzioni ben al di sotto del Limite di Onsager [35]. Le fasi LC esibite dall'nDNA sono le stesse osservate nel DNA lungo (N^* e COL esagonale) sebbene si trovino a concentrazioni più elevate, ad esempio intorno a 750 mg/mL per l' N^* . La formazione di fasi LC di nDNA è una conseguenza delle interazioni di accoppiamento e impilamento.

La prima osservazione di LC nell'nDNA ha comportato l'uso di sequenze auto-complementari. Quando questi tipi di oligomeri si accoppiano, formano doppie eliche. Le basi terminali di diversi duplex tendono a impilarsi per evitare l'esposizione all'acqua, portando a un'interazione attrattiva che li tiene insieme quando le loro facce piatte aromatiche sono in contatto. Queste interazioni end-to-end inducono un'aggregazione lineare reversibile di oligonucleotidi in catene lunghe.

Quando le catene legate raggiungono un rapporto assiale e una concentrazione sufficientemente grandi da entrare nel regime di Onsager, la soluzione si ordina nella fase N^* e, aumentando ulteriormente la concentrazione, nella fase COL (figura 2.18). La variazione di energia libera associata allo stacking dei duplex, stimata dal diagramma di fase nDNA, $\Delta G_{BS} \sim 4 - 6k_B T$, è più grande, ma simile, a quella prevista per lo stacking di base all'interno di eliche [60]. La libertà dei duplex impilati di ruotare attorno al proprio asse può consentire al sistema di trovare una condizione di impilamento migliore (superficie impilata più ampia) rispetto ai normali quadrupletti all'interno di doppie eliche. Questo fatto potrebbe effettivamente spiegare

perché l'energia libera coinvolta nei duplex impilati supera quella per l'impilamento all'interno di eliche.

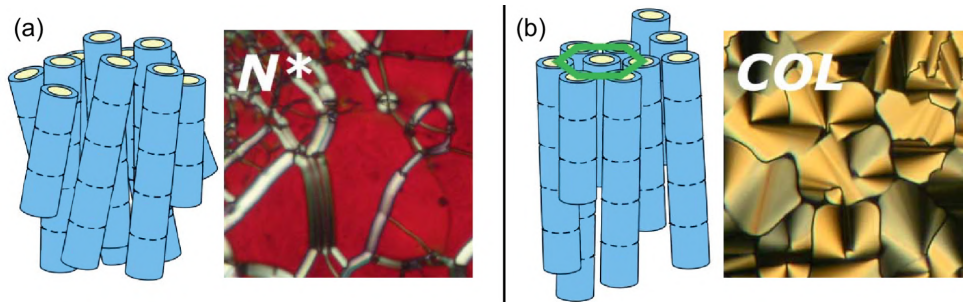


Figura 2.18. Si mostrano le fasi N^* (a) e COL LC (b) possibili con le eliche capaci di aggregazione lineare. Immagine tratta da [61].

È stato scoperto che lo stacking di nucleobasi è coinvolto in diversi processi di autoassemblaggio del DNA, promuovendo la formazione di complessi DNA-DNA e DNA-proteina in vivo [44], guidando interazioni end-to-end di DNA double e triple-eliche in soluzioni semidiluite [29], o determinare la geometria dei cristalli di oligonucleotidi di DNA [59].

Capitolo 3

Simulazioni Monte Carlo

In questo capitolo viene descritto il metodo Monte Carlo e si fornisce una breve descrizione della meccanica statistica degli oggetti fisici di interesse nella tesi. Per effettuare lo studio della materia soffice e in particolare delle fasi liquido-cristalline si devono calcolare i valor medi degli osservabili per sistemi con un gran numero di particelle, nell'ordine delle migliaia. Per valutare i valor medi è necessario calcolare gli integrali su tutti i gradi di libertà, nello spazio delle fasi, delle particelle. Si consideri il caso in cui si voglia calcolare il valor medio dell'osservabile generico A :

$$\langle A(\mathbf{r}_N) \rangle = \frac{1}{Z} \int d\mathbf{r}^N d\mathbf{p}^N e^{-\beta \mathbf{H}(\mathbf{r}_N, \mathbf{p}_N)} A(\mathbf{r}^N) \quad (3.1)$$

Dove Z è la funzione di partizione, definita da:

$$Z \equiv \int d\mathbf{r}^N d\mathbf{p}^N e^{-\beta \mathbf{H}(\mathbf{r}_N, \mathbf{p}_N)} \quad (3.2)$$

Dove $\mathbf{H}(\mathbf{r}_N, \mathbf{p}_N) = U(\mathbf{r}_N) + K(\mathbf{p}_N)$ ovvero la somma dell'energia di legame delle particelle e dell'energia cinetica.

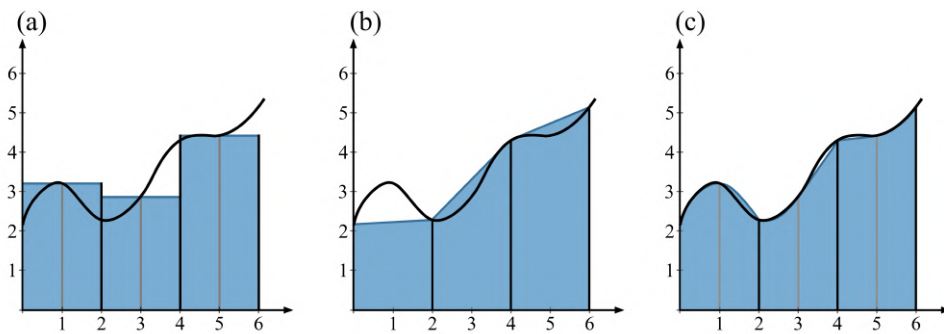


Figura 3.1. Rappresentazione grafica dei metodi numerici: (a) Metodo dei rettangoli. (b) Metodo dei trapezi. (c) Metodo Cavalieri-Simpson, dove si approssima la funzione con una parabola per ogni zona.

Il calcolo dell'integrale nell'equazione (3.1) può essere fatto utilizzando diversi metodi numerici, l'idea generale è quella di discretizzare l'intervallo di integrazione, come negli integrali di Riemann, e di moltiplicare ogni segmento per il valore della funzione

in quella zona. Con questa tecnica è possibile sviluppare diversi metodi tra cui: il metodo dei rettangoli, il metodo dei trapezi e il metodo di Cavalieri-Simpson (immagine 3.1).

Tuttavia, è facile notare che utilizzare uno di questi metodi per sistemi con un grande numero di coordinate non è conveniente. Si supponga di voler utilizzare la quadratura numerica per valutare un integrale su $N = 50$ particelle in uno spazio tridimensionale, ovvero $6N$ gradi di libertà, e di valutare l'integrale su $m = 10$ punti nel dominio equidistanti. Di conseguenza il numero di punti totali su cui si integrerà sarà pari a $m^{6N} = 10^{300}$. Il supercomputer ad ora più potente¹ è in grado di effettuare 10^{17} operazioni al secondo, di conseguenza per effettuare l'integrazione sarebbero necessari 10^{283} secondi, che è di gran lunga più grande dell'età dell'universo (dell'ordine di 10^{17} secondi).

Questo esempio dimostra che è necessario trovare tecniche migliori per calcolare le medie termiche. Una di queste tecniche è il metodo Monte Carlo o, più precisamente, l'algoritmo importance-sampling Monte Carlo introdotto nel 1953 [33].

3.1 Importance Sampling

Prima di discutere l'importance sampling, si discute una tecnica Monte Carlo semplice, ovvero il random sampling (detto anche Monte Carlo *hit-and-miss*). Si supponga che si voglia calcolare numericamente l'integrale unidimensionale:

$$\mathbf{I} = \int_a^b \mathbf{f}(x) dx \quad (3.3)$$

Invece di utilizzare una quadratura convenzionale dove l'integrando è valutato a valori di ascissa determinate si procede a riscrivere l'ultima equazione come:

$$\mathbf{I} = (b - a) \langle \mathbf{f}(x) \rangle \quad (3.4)$$

dove $\langle \mathbf{f}(x) \rangle$ è la media, non pesata, di $\mathbf{f}(x)$ sull'intervallo $[a, b]$. Nel caso del Monte Carlo *brute force* questa media è valutata calcolando $\mathbf{f}(x)$ per un gran numero L di valori di x randomicamente distribuiti sull'intervallo $[a, b]$. Nel limite in cui $L \rightarrow \infty$ questa procedura restituisce il valore corretto di \mathbf{I} . Questo metodo non è ottimale siccome si perde tempo computazionale a generare punti dove il fattore di Boltzmann è trascurabile nel caso di medie statistiche. Chiaramente sarebbe preferibile campionare più densamente zone dove il fattore di Boltzmann è grande, questa è l'idea che c'è dietro all'importance sampling.

Con importance sampling si intende un processo di integrazione Monte Carlo nella quale i punti non vengono scelti in maniera random da una distribuzione uniforme ma attraverso una determinata distribuzione $w(x)$. Per capire quale sia una buona scelta di questa distribuzione si consideri un caso semplice unidimensionale, si cerca

¹Il supercomputer Frontier costruito da Hewlett Packard Enterprise (HPE) si trova al laboratorio nazionale di Oak Ridge (ORNL) in Tennessee, USA.

sempre di valutare l'integrale (3.3), ma con i punti di campionamento distribuiti non uniformemente con distribuzione $w(x)$ sull'intervallo $[a, b]$ (dove si impone, per convenienza, $a = 0$ e $b = 1$). Di conseguenza è possibile riscrivere l'integrale come:

$$\mathbf{I} = \int_0^1 dx w(x) \frac{\mathbf{f}(x)}{w(x)} \quad (3.5)$$

Si assuma che $w(x)$ derivi da un'altra funzione $u(x)$, non-negativa e non-decrescente, con $u(0) = 0$ e $u(1) = 1$. Di conseguenza $u(x)$ è normalizzata. Sostituendola:

$$\mathbf{I} = \int_0^1 du \frac{\mathbf{f}[x(u)]}{w[x(u)]} \quad (3.6)$$

Dove con $x(u)$ si indica che se u è la variabile di integrazione allora x verrà espressa in funzione di u . Successivamente si generano L valori random di u uniformemente distribuiti nell'intervallo $[0, 1]$. Di conseguenza l'integrale sarà pari a:

$$\mathbf{I} \approx \frac{1}{L} \sum_{i=1}^L \frac{\mathbf{f}[x(u_i)]}{w[x(u_i)]} \quad (3.7)$$

Per capire il vantaggio che ne consegue da questa scrittura di \mathbf{I} si calcola la deviazione standard σ_I^2 di I_L , dove quest'ultima è ricavata dall'equazione (3.7) con L punti random di campionamento:

$$\sigma_I^2 = \frac{1}{L^2} \sum_{i=1}^L \sum_{j=1}^L \left\langle \left(\frac{\mathbf{f}[x(u_i)]}{w[x(u_i)]} - \langle \mathbf{f}/w \rangle \right) \left(\frac{\mathbf{f}[x(u_j)]}{w[x(u_j)]} - \langle \mathbf{f}/w \rangle \right) \right\rangle \quad (3.8)$$

Dove le parentesi angolari indicano la vera media, quella ottenuta nel limite in cui $L \rightarrow \infty$. Se si assumono i campionamenti i e j completamente indipendenti i termini misti si cancellano e rimane solamente:

$$\sigma_I^2 = \frac{1}{L^2} \sum_{i=1}^L \left\langle \left(\frac{\mathbf{f}[x(u_i)]}{w[x(u_i)]} - \langle \mathbf{f}/w \rangle \right)^2 \right\rangle = \frac{1}{L} [\langle (\mathbf{f}/w)^2 \rangle - \langle \mathbf{f}/w \rangle^2] \quad (3.9)$$

L'equazione (3.9) mostra che la varianza in \mathbf{I} ha un andamento come $1/L$, ma la grandezza di questa varianza è possibile controllarla scegliendo $w(x)$ tale per cui $\mathbf{f}(x)/w(x)$ sia una funzione liscia in x ovvero il più possibile costante. Si noti inoltre che nel caso $w(x)$ sia costante, come nel campionamento brute force Monte Carlo, l'errore in I va come $1/\sqrt{L}$.

Quindi, per avere un buon campionamento, si sceglie una distribuzione di probabilità $w(x)$ simile, se non proporzionale, alla distribuzione di equilibrio. Ad esempio, per effettuare un campionamento nell'ensemble canonico, sarà necessario conoscere la distribuzione $e^{-\beta U(r_i)}/Z$ e di conseguenza la funzione di partizione che però è l'obiettivo che si cercava di raggiungere con l'importance sampling. Tuttavia come vedremo a breve per generare dei punti con la giusta distribuzione non è necessario calcolare Z .

3.2 Il metodo Metropolis

Nella maggior parte delle applicazioni, l'obiettivo delle simulazioni molecolari Monte Carlo è di calcolare delle medie d'*ensemble* di osservabili tramite i quali poter determinare il diagramma di fase del sistema. Per iniziare si consideri un ensemble canonico con un numero di atomi N , il volume e la temperatura costanti. Si supponga si voglia calcolare la media termica $\langle A \rangle$ di un osservabile A , per farlo si valuta la media dell'ensemble utilizzando la formula:

$$\langle A(\mathbf{r}^N) \rangle = \frac{\int d\mathbf{r}^N e^{-\beta U(\mathbf{r}^N)} A(\mathbf{r}^N)}{\int d\mathbf{r}^N e^{-\beta U(\mathbf{r}^N)}} \quad (3.10)$$

Dove \mathbf{r}^N indica le coordinate delle N particelle, $\beta = 1/k_B T$, U è l'energia potenziale del sistema. Inoltre il termine $e^{-\beta U}$ viene chiamato il fattore di Boltzmann. Quindi si cerca di calcolare il rapporto dei due integrali. Il metodo Metropolis mostra che è possibile scegliere uno schema Monte Carlo per calcolare il rapporto degli integrali. Si definisce la funzione di partizione come il denominatore della formula (3.10):

$$Z \equiv \int d\mathbf{r}^N e^{-\beta U(\mathbf{r}^N)} \quad (3.11)$$

Si definisce quindi la densità di probabilità di trovare il sistema nella configurazione \mathbf{r}^N come:

$$\mathcal{N}(\mathbf{r}^N) \equiv \frac{e^{-\beta U(\mathbf{r}^N)}}{Z} \quad (3.12)$$

Si consideri ora come generare una configurazione di punti con relativa probabilità pari al fattore di Boltzmann. Per farlo, si utilizzano delle opportune catene di Markov, descritte nel prossimo paragrafo, che si possono ottenere generando una configurazione di prova e accettandola con un'opportuna probabilità.

3.2.1 Catene di Markov

La funzione di partizione Z non è calcolabile direttamente di conseguenza non è possibile calcolare la distribuzione di Boltzmann. Per calcolare quest'ultima si possono però utilizzare le catene di Markov. Un processo stocastico Markoviano è definito come un processo aleatorio in cui la probabilità di transizione che determina il passaggio a uno stato di sistema dipende solamente dallo stato del sistema immediatamente precedente. Le catene di Markov sono definite dalle tre seguenti proprietà:

1. $\{x_t\}_{t=1,\dots,L} = \{x_1, \dots, x_L\}$ con $x_t \in S$ dove S è lo spazio degli stati finito.
2. $P(x_{t+1} = p_{t+1} \mid x_t = p_t, \dots, x_1 = p_1) = P(x_{t+1} = p_{t+1} \mid x_t = p_t)$ si assume che il valore di x_{t+1} dipenda solo dalla variabile precedente x_t .
3. La catena è ergodica ovvero tutti i punti dello spazio delle fasi sono raggiungibili in un numero finito di transizioni.

Definita la probabilità di transizione da $n \rightarrow k$ come $\Pi_{nk} = P(x_{t+1} = p_k | x_t = p_n)$ è possibile definire come evolve la distribuzione iniziale di probabilità:

$$p_n^{t+1} = p_n^t - \sum_k p_n^t \Pi_{nk} + \sum_k p_k^t \Pi_{kn} \quad (3.13)$$

Dove con Π_{nk} si indicano le simulazioni che vanno da $n \rightarrow k$ e con Π_{kn} il processo inverso. Nel limite di t grande, dalle proprietà delle catene di Markov, si avrà che $p_n^{t+1} = p_n^t = p_n$. Di conseguenza, semplificando nell'equazione 3.13, si ottiene:

$$-\sum_k p_n^t \Pi_{nk} + \sum_k p_k^t \Pi_{kn} = -p_n \sum_k \Pi_{nk} + \sum_k p_k \Pi_{kn} = 0 \quad (3.14)$$

Infine visto che per la normalizzazione vale che $\sum_k \Pi_{nk} = 1$ si ottiene la condizione di equilibrio:

$$p_n = \sum_k p_k \Pi_{kn} \quad (3.15)$$

Questa condizione afferma che si è raggiunto l'equilibrio nel momento in cui la probabilità che il sistema lasci lo stato deve essere uguale alla probabilità che il sistema arrivi in quello stato dagli altri stati.

3.2.2 Bilancio dettagliato

Si discute quindi lo schema Metropolis, questo permetterà il calcolo della probabilità di transizione $\pi(o \rightarrow n)$ dalla configurazione o alla n . La probabilità $\pi(o \rightarrow n)$ deve sottostare alla condizione per cui non distrugga una distribuzione di equilibrio una volta che questa sia raggiunta. Di conseguenza deve rispettare la condizione di equilibrio definita nell'equazione 3.15. Si può imporre una condizione più stringente per convenienza: all'equilibrio il numero medio delle mosse accettate dallo stato o a un qualsiasi altro stato n è pari al numero di mosse nel verso opposto. Questa condizione viene chiamata bilancio dettagliato, che formalmente può essere espressa come:

$$\mathcal{N}(o) \cdot \pi(o \rightarrow n) = \mathcal{N}(n) \cdot \pi(n \rightarrow o) \quad (3.16)$$

Si divide quindi la mossa Monte Carlo in due passi. La matrice di transizione che determina la probabilità di fare un tentativo di mossa da o a n chiamata $\alpha(o \rightarrow n)$. Successivamente si decide se accettare o rifiutare la mossa, la probabilità di accettare la mossa viene denominata $acc(o \rightarrow n)$. Di conseguenza:

$$\pi(o \rightarrow n) = \alpha(o \rightarrow n) \cdot acc(o \rightarrow n) \quad (3.17)$$

Nello schema Metropolis originale, α è scelto come una matrice simmetrica, di conseguenza: $\alpha(o \rightarrow n) = \alpha(n \rightarrow o)$. In generale è possibile costruire schemi anche non simmetrici. Tuttavia supponendo che la matrice sia simmetrica è possibile riscrivere l'equazione (3.16) come:

$$\mathcal{N}(o) \cdot acc(o \rightarrow n) = \mathcal{N}(n) \cdot acc(n \rightarrow o) \quad (3.18)$$

Di conseguenza si può riscrivere come:

$$\frac{acc(o \rightarrow n)}{acc(n \rightarrow o)} = \frac{\mathcal{N}(n)}{\mathcal{N}(o)} = e^{-\beta[\mathcal{U}(n) - \mathcal{U}(o)]} \quad (3.19)$$

Esistono varie scelte per $acc(o \rightarrow n)$ in tutte però questa quantità non può essere maggiore di 1. La scelta nello schema Metropolis è che:

$$\begin{aligned} acc(o \rightarrow n) &= \frac{\mathcal{N}(n)}{\mathcal{N}(o)} & \text{se} & \quad \mathcal{N}(n) < \mathcal{N}(o) \\ acc(o \rightarrow n) &= 1 & \text{se} & \quad \mathcal{N}(n) \geq \mathcal{N}(o) \end{aligned} \quad (3.20)$$

Le altre scelte possibili per $acc(o \rightarrow n)$ risultano meno efficienti nell'esplorare lo spazio delle configurazioni rispetto a quella proposta dallo schema Metropolis. Inoltre si noti che non ci sono vincoli sulla matrice acc se non che debba essere simmetrica, questo si riflette in una grande libertà nella scelta del tentativo di mossa. Riassumendo, una volta generato il tentativo di mossa dallo stato o a n con $\mathcal{U}(n) < \mathcal{U}(o)$, la probabilità di accettazione della mossa è data da:

$$acc(o \rightarrow n) = e^{-\beta[\mathcal{U}(n) - \mathcal{U}(o)]} < 1 \quad (3.21)$$

Si genera quindi un numero casuale $Rand$ con distribuzione uniforme nell'intervallo tra $[0,1]$ e si accetta la mossa se $Rand < acc(o \rightarrow n)$ e rifiutarla nel caso opposto.

3.2.3 Codice

Per eseguire una simulazione utilizzando il metodo Metropolis si utilizza il seguente schema:

1. Si seleziona in modo random una particella e si calcola la sua energia $\mathcal{U}(o)$.
2. Si sposta la particella selezionata di una quantità random, $\mathbf{r}' = \mathbf{r} + \Delta$ e si calcola la sua nuova energia $\mathcal{U}(n)$.
3. Si accetta la mossa da o a n con probabilità:

$$acc(o \rightarrow n) = \min(1, \exp[-\beta[\mathcal{U}(n) - \mathcal{U}(o)]]$$

Lo spostamento Δ può essere effettuato aggiungendo a ogni coordinata x , y e z una quantità random tra $-\Delta/2$ e $\Delta/2$. Si nota che la mossa che va da \mathbf{r} a \mathbf{r}' è ugualmente probabile alla mossa che va da \mathbf{r}' a \mathbf{r} , infatti α è simmetrica.

Per l'economia del programma è importante scegliere un valore di Δ adatto, se questo fosse troppo grande la configurazione finale avrebbe un'energia elevata e la mossa sarebbe probabilmente rifiutata. All'opposto, se fosse uno spostamento piccolo, il cambiamento nell'energia sarà probabilmente piccolo e la mossa sarà accettata. Si deve quindi valutare il rapporto tra le mosse accettate e tutte le mosse tentate, non esiste un valore ottimale di questo rapporto ma in generale si cerca di trovare un valore che permetta di esplorare tutto lo spazio delle configurazioni e tipicamente questo valore deve essere compreso tra 0.3 e 0.5.

3.3 Ensembles

Esistono diversi *ensembles* che vengono spesso utilizzati per effettuare le simulazioni Monte Carlo. L'insieme canonico (NVT) [37] si riferisce a tutti gli stati coerenti con temperatura, volume e numero di particelle fissati. L'*Ensemble* Gran Canonico (μ VT) [1] si riferisce al sistema con volume, temperatura e potenziale chimico fissi, che può essere utilizzato per calcolare il numero medio di molecole negli studi di assorbimento. L'insieme isobarico-isotermico (NPT) è vincolato da pressione, temperatura e numero costante di molecole. L'insieme Gibbs-NVT tiene conto dell'equilibrio di fase con temperatura, volume e numero di particelle costanti. L'ensemble Gibbs-NPT si occupa di sistemi a temperatura, pressione e numero di molecole costanti per l'equilibrio di fase. In questo capitolo verranno introdotti l'ensemble NVT e l'ensemble NPT, che verranno utilizzati per le simulazioni Monte Carlo nei prossimi capitoli.

3.4 NVT

Nell'*ensemble* canonico il numero di particelle, la temperatura e il volume sono mantenute costanti. Nella mossa Monte Carlo si effettua uno spostamento delle particelle.

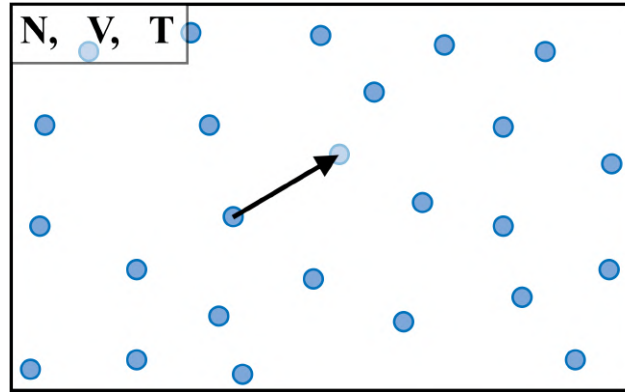


Figura 3.2. Schema di un *Ensemble* canonico con N , V e T costanti.

La funzione di partizione, di questo sistema, sarà data da:

$$Q(N, V, T) \equiv \frac{1}{\Lambda^{3N} N!} \int d\mathbf{r}^N \exp[-\beta \mathcal{U}(\mathbf{r}^N)] \quad (3.22)$$

dove $\Lambda = \sqrt{h^2/(2\pi m k_B T)}$ è la lunghezza d'onda di De Broglie. La probabilità di trovare il sistema nella configurazione \mathbf{r}^N è data da:

$$\mathcal{N}(\mathbf{r}^N) \propto \exp[-\beta \mathcal{U}(\mathbf{r}^N)] \quad (3.23)$$

Le equazioni (3.22) e (3.23) sono la base per effettuare una simulazione nell'*ensemble* canonico.

3.4.1 Implementazione codice

Nell'*ensembles* NVT è possibile realizzare due tipi di mosse, quella traslazionale e quella rotazionale. La seconda è disponibile solo nel caso in cui non si abbiano particelle sferiche e l'algoritmo di rotazione dipende dalle simmetrie delle particelle utilizzate.

Mossa di traslazione

Per effettuare una mossa traslazionale Monte Carlo nell'*ensembles* NVT si utilizza il seguente schema:

1. Si seleziona in modo random una particella e si calcola la sua energia $\mathcal{U}(o)$.
2. Si sposta la particella selezionata di una quantità random

$$\mathbf{r}(o) \rightarrow \mathbf{r}(o) + \Delta(\text{Rand} - 0.5)$$

dove $\Delta/2$ è lo spostamento massimo e va ottimizzato in base alla simulazione eseguita. La nuova configurazione è denominata n e avrà energia $\mathcal{U}(n)$.

3. Si accetta la mossa da o a n con probabilità:

$$\text{acc}(o \rightarrow n) = \min(1, \exp[-\beta(\mathcal{U}(n) - \mathcal{U}(o))])$$

Se rifiutata, la vecchia configurazione è mantenuta.

Mossa di rotazione

Lo schema, per effettuare la mossa di rotazione, è analogo a quello della mossa di traslazione. Tuttavia al punto 2 si procede a ruotare la particella. Per effettuare una rotazione esistono diversi algoritmi e la scelta dipende dal tipo di particella in esame. Si presenta ora un metodo per ruotare una particella descritta dai tre versori, nel sistema di riferimento del laboratorio, \mathbf{u}_i con $i = 1, 2, 3$. Si genera un vettore casuale sulla sfera unitaria e si procede a ruotare la particella intorno a quell'asse di un angolo pari a $\xi\Delta\theta$, dove $\xi \in [0, 1]$ è un numero generato casualmente e $\Delta\theta$ è l'angolo di rotazione massimo, intorno a l'asse generato, che andrà ottimizzato per velocizzare la simulazione. La matrice di rotazione \mathbf{R} è costruita come:

$$\mathbf{R} = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)^T = \begin{bmatrix} u_{1x} & u_{1y} & u_{1z} \\ u_{2x} & u_{2y} & u_{2z} \\ u_{3x} & u_{3y} & u_{3z} \end{bmatrix} \quad (3.24)$$

Per generare un vettore casuale sulla sfera unitaria si utilizza l'algoritmo di Marsaglia, che consiste nella generazione di due numeri randomici $\xi_1, \xi_2 \in [-1, 1]$ finchè non soddisfano la condizione $\xi_1^2 + \xi_2^2 < 1$. Il vettore casuale generato avrà quindi componenti:

$$\hat{\mathbf{v}} = (2\xi_1(1 - \xi_1^2 + \xi_2^2)^{1/2}, 2\xi_2(1 - \xi_1^2 + \xi_2^2)^{1/2}, 1 - 2(\xi_1^2 + \xi_2^2)) \quad (3.25)$$

Quindi viene generata la matrice $\mathbf{\Omega}$:

$$\mathbf{\Omega} = \begin{bmatrix} 0 & -v_z & v_y \\ v_z & 0 & -v_x \\ -v_y & v_x & 0 \end{bmatrix} \quad (3.26)$$

Viene generato casualmente l'angolo di rotazione $\theta \in [-\Delta\theta, \Delta\theta]$ e calcolata la matrice \mathbf{M} , definita come:

$$\mathbf{M} = -\sin(\theta)\mathbf{\Omega} + (1 - \cos(\theta))\mathbf{\Omega}^2 \quad (3.27)$$

Viene infine aggiornata la matrice di rotazione $\mathbf{R}' = \mathbf{R}(\mathbf{I} + \mathbf{M})$, dove \mathbf{I} è la matrice identità.

3.5 NPT

L'ensemble isotermico-isobarico (NPT) è frequentemente usato nelle simulazioni Monte Carlo, questo perché molti esperimenti effettuati nei laboratori vengono eseguiti sotto condizioni di pressione e temperatura controllate. Questo ensemble è conveniente anche nel momento in cui si è nelle vicinanze di una transizione di fase del primo ordine, vista la pressione costante il sistema non potrà trovarsi in condizioni di coesistenza poiché dovrà trovarsi in un minimo dell'energia libera di Gibbs. Al contrario nelle simulazioni NVT il sistema potrebbe essere mantenuto a una densità dove forma due zone di differenti fasi a differenti densità.

3.5.1 Basi di Meccanica Statistica

Si assuma, per semplicità, di aver un sistema composto da N atomi identici. La funzione di partizione di questo sistema è data da:

$$Q(N, V, T) = \frac{1}{\Lambda^{3N} N!} \int_0^L \dots \int_0^L e^{-\beta \mathcal{U}(\mathbf{r}^N)} d\mathbf{r}^N \quad (3.28)$$

Conviene riformulare questa equazione, si assume che il sistema sia contenuto in una box di volume $V = L^3$ e si definiscono le coordinate scalate \mathbf{s}^N dove $\mathbf{r}_i = L\mathbf{s}_i$.

Se sostituiamo le coordinate scalate nell'espressione della funzione di partizione, si ottiene:

$$Q(N, V, T) = \frac{V^N}{\Lambda^{3N} N!} \int_0^1 \dots \int_0^1 e^{-\beta \mathcal{U}(\mathbf{s}^N, L)} d\mathbf{s}^N \quad (3.29)$$

Dove si usa $\mathcal{U}(\mathbf{s}^N, L)$ per indicare che \mathcal{U} dipenda dalle distanze reali tra le particelle e non scalate. L'espressione per l'energia libera di Helmholtz del sistema è pari a:

$$\begin{aligned} F(N, V, T) &= -k_B T \ln Q(N, V, T) = \\ &= -k_B T \ln \frac{V^N}{\Lambda^{3N} N!} - k_B T \ln \int_0^1 \dots \int_0^1 e^{-\beta \mathcal{U}(\mathbf{s}^N, L)} d\mathbf{s}^N = \\ &= F^{idealgas}(N, V, T) + F^{excess}(N, V, T) \end{aligned} \quad (3.30)$$

Si è divisa l'energia libera di Helmholtz in due contributi, una è la parte di gas ideale e, l'altra, una parte in eccesso. Si assuma, come mostrato in figura (3.3), che

il sistema sia immerso in un bagno termico di un gas ideale e sia separato da esso da un pistone che mantiene entrambi i sistemi a pressione P . Si assuma inoltre che le il sistema sia composto in totale da M particelle in un volume totale pari a V_0 come mostrato in figura (3.3).

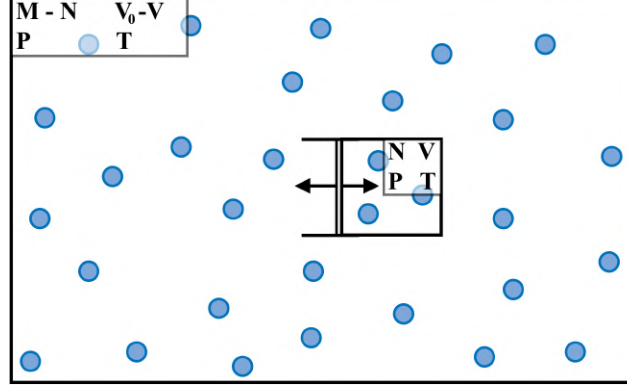


Figura 3.3. Un gas ideale composto da M particelle e volume $V_0 - V$ può scambiare volume con un sistema composto da N particelle.

Siano N e V rispettivamente il numero di particelle e il volume del box di cui si vuole calcolare l'energia libera di Helmholtz si assume inoltre che $M \gg N$ e $V_0 \gg V$. La funzione di partizione di tutto il sistema è il prodotto della funzione di partizione del sottosistema S con volume V definita da:

$$Q_S(N, V, T) = \frac{V^N}{\Lambda^{3N} N!} \int d\mathbf{s}^N e^{-\beta \mathcal{U}(\mathbf{s}^N, L)} \quad (3.31)$$

Mentre il sistema serbatoio R avrà funzione di partizione pari a:

$$Q_R(M - N, V_0 - V, T) = \frac{(V_0 - V)^{M-N}}{\Lambda^{3(M-N)} (M - N)!} \int d\mathbf{s}^{M-N} \quad (3.32)$$

Dove il termine di Boltzmann non è stato inserito perché questa è la funzione di partizione di un gas ideale. La funzione di partizione totale sarà quindi:

$$Q(N, M, V, V_0, T) = Q_S(N, V, T) \cdot Q_R(M - N, V_0 - V, T) \quad (3.33)$$

Per compattezza si è assunto che la lunghezza d'onda di De Broglie delle particelle del gas ideale sia sempre Λ . L'energia libera del sistema sarà data dall'equazione (3.30) utilizzando la funzione di partizione totale.

Si consideri ora il limite in cui la grandezza del serbatoio tendi a infinito ($V_0 \rightarrow \infty, M \rightarrow \infty, (M - V)/V_0 \rightarrow \rho$), in questo caso un cambio di volume nel sottosistema non cambia la pressione P del serbatoio, che funzionerà come un pressostato. Ricordando le relazioni termodinamiche:

$$F = U - TS \quad G = N\mu = F + PV \quad \beta F = N\beta\mu - \beta PV \quad (3.34)$$

e la relazione:

$$\beta F_R = -\ln[Q_R(M - N, V_0 - V, T)] \quad (3.35)$$

quindi si può riscrivere come:

$$Q_R(M - N, V_0 - V, T) = e^{-\beta F_R(M-N, V_0-V, T)} = e^{-(M-N)\beta\mu + \beta P(V_0-V)} \quad (3.36)$$

Si noti che la dipendenza dal sottosistema è solo nel termine $[-\beta PV]$. Si supponga ora che il pistone sia libero di muoversi. Di conseguenza è necessario considerare nella funzione di partizione nell'equazione (3.31) tutti i possibili valori di V e sommare su di essi.

$$\begin{aligned} Q(N, P, T) &= \beta P \int_0^\infty dV e^{-(M-N)\beta\mu} e^{\beta P(V_0-V)} Q_S(N, V, T) = \\ &= \beta P e^{-(M-N)\beta\mu} e^{\beta P V_0} \int_0^\infty dV e^{\beta P V} Q_S(N, V, T) \end{aligned} \quad (3.37)$$

Sostituendo Q_S e considerando che N, P, T e M sono quantità costanti possono essere accorpate al termine di normalizzazione.

$$Q(N, P, T) = \beta P \int dV V^N e^{-\beta P V} \int ds^N e^{-\beta \mathcal{U}(s^N, L)} \quad (3.38)$$

I termini βP sono stati inclusi per rendere la funzione di partizione adimensionale. Il valore più probabile di V sarà quello che minimizzerà l'energia libera del sistema complessivo. La densità di probabilità $\mathcal{N}(V)$ che il sottosistema di N particelle abbia volume V è data da:

$$\mathcal{N}_{N,P,T}(V) = \frac{V^N e^{-\beta P V} \int ds^N e^{-\beta \mathcal{U}(s^N, L)}}{\int dV' V'^N e^{-\beta P V'} \int ds^N e^{-\beta \mathcal{U}(s^N, L')}} \quad (3.39)$$

La densità di probabilità di trovare il sottosistema nella configurazione con N atomi a un dato volume V è riscrivibile come:

$$\mathcal{N}(V, s^N) \approx V^N e^{-\beta P V} \int ds^N e^{-\beta \mathcal{U}(s^N, L)} = e^{-\beta \mathcal{U}(s^N, L) + P V - N \beta^{-1} \ln V} \quad (3.40)$$

É ora possibile effettuare la simulazione sulle coordinate s^N e il volume V . Utilizzando il metodo Monte Carlo, V è trattato come una coordinata addizionale e i tentativi di mossa su V soddisfano le stesse regole delle mosse su s . Si consideri un tentativo di mossa che cambi il volume da V a $V' = V + \Delta V$ dove ΔV è un numero random generato uniformemente nell'intervallo $[-\Delta V_{max}, \Delta V_{max}]$. Nello schema Metropolis la mossa sarà accettata con probabilità:

$$acc(o \rightarrow n) = \min(1, e^{-\beta(\mathcal{U}(s^N, V') - \mathcal{U}(s^N, V)) + P(V' - V) - N \beta^{-1} \ln(V'/V)}) \quad (3.41)$$

3.5.2 Implementazione Codice

Il costo computazionale di effettuare un cambio di volume è maggiore rispetto a una mossa di spostamento della particella. Un tentativo di mossa sul volume richiede di ricalcolare tutte le interazioni tra le particelle, di conseguenza ha un costo pari a circa N tentativi di mossa su le posizioni delle particelle. Si effettua quindi un tentativo di mossa sul volume ogni N tentativi sulla posizione, ma per mantenere la simmetria della catena di Markov, il cambio di volume non deve essere effettuato periodicamente ma ad ogni passo ci sarà una probabilità pari a $1/N$.

Mossa di cambio volume

Lo schema per la mossa di variazione del volume è diviso nel seguente modo: si modifica il volume di una quantità scelta casualmente dentro un certo range, si procede a scalare la posizione di tutte le particelle e a controllare se il cambio volume abbia creato intersezioni tra particelle. In caso positivo la mossa viene rifiutata. Si procede calcolando l'energia della nuova configurazione e ad accettare la mossa con probabilità descritta dall'equazione: 3.41. Se nel sistema studiato è presente un raggio di cutoff quest'ultimo andrà scalato quando varierà la dimensione della box, varierà con L : $r'_C = (V'/V)r_C$. Va posta anche attenzione se nel sistema vengono utilizzate le *linked cell list*, discusse tra qualche paragrafo, infatti le dimensioni delle celle devono essere fissate in base alla dimensioni degli oggetti studiati di conseguenza se il sistema venisse compresso andrebbero diminuito il numero delle celle. Nel caso contrario, per ottimizzare il tempo computazionale, è necessario aumentare il numero delle celle.

Per una ulteriore ottimizzazione è possibili utilizzare due tipi di espansione del volume, isotropa e anisotropa.

1. **Espansione isotropa:** in questo caso la scatola della simulazione rimane della stessa forma iniziale, infatti, se cubica, questa rimarrà tale andando a cambiare ogni lato della stessa quantità. Quindi se lo stato iniziale ha volume V verrà generato un numero casuale $Rand \in [0, 1]$ e il volume del passo successivo sarà: $V' = V + \Delta V(Rand - 0.5)$. Successivamente andranno anche scalate tutte le posizioni delle particelle che avranno posizione $\mathbf{r}' = \mathbf{r}(V'/V)^{1/3}$.
2. **Espansione anisotropa:** nel caso in cui la fase finale di equilibrio possa avere una struttura sfavorita da una scatola cubica, ad esempio nella fase colonnare, è possibile andare a realizzare una mossa di cambio volume andando a modificare solo la dimensione di un lato della box. Il lato modificato verrà scelto casualmente a ogni mossa di cambio volume, si procede quindi a generare un numero casuale $Rand \in [0, 1]$ e a calcolare $L'_i = L_i + \Delta L(Rand - 0.5)$. In questo caso la posizione delle particelle andrà aggiornata come $r'_i = r_i(L'_i/L_i)$.

Capitolo 4

Metodi: Approccio computazionale e design delle simulazioni

In questo capitolo si presentano le ottimizzazioni e i metodi utilizzati nel codice scritto per le simulazioni eseguite nei prossimi capitoli. Nella prima parte si descrivono i due tipi di particelle utilizzate in questa tesi, cilindri e dischi tagliati longitudinalmente e si giustifica la scelta di questi due oggetti. Vengono anche elencati i due algoritmi utilizzati per trovare l'intersezione tra poliedri, con una focalizzazione all'ottimizzazione del codice per ridurre il tempo per completare una simulazione. Si procede quindi al confronto dei tempi di esecuzione per una simulazione *NVT* utilizzando i due algoritmi di intersezione di poliedri.

4.1 Presentazione dei due modelli

Il programma realizzato permette di simulare sistemi costituiti da oggetti di forma arbitraria negli *ensemble NVT* e *NPT*, con l'aggiunta di siti attrattivi. In questa tesi l'attenzione sarà rivolta a simulare dei cilindri e dei dischetti tagliati longitudinalmente. Un primo obiettivo sarà quello di simulare filamenti di DNA di lunghezza pari a 12 basi, quindi per un totale quindi di 24 basi e circa 4 nanometri di lunghezza, approssimandoli a dei cilindri di *aspect-ratio* $X_0 = h/D = 2$ con all'estremità una o due patch attrattive per simulare la possibilità dei filamenti di DNA di accoppiarsi e formare catene più lunghe ovvero per modellizzare le interazioni di *stacking*. Successivamente verrà studiato un modello che approssima la singola base come un dischetto tagliato longitudinalmente, inserendo un totale di quattro patch, due per simulare il fenomeno di *stacking* delle basi e due per simulare il fenomeno di *pairing*.

Nel momento in cui si procede a effettuare un taglio nella forma di un poliedro regolare non è più possibile utilizzare gli algoritmi di intersezione classici per quel determinato poliedro. Di conseguenza è necessario trovare un metodo per considerare anche le intersezioni miste del tipo taglio-superficie curva del cilindro, in genere questo è un problema complesso e spesso la soluzione rimane strettamente legata alla struttura della particella scelta, modificando questa il programma andrebbe

riadattato. Per questo si è scelto di utilizzare un approccio più generale, ovvero si approssima con un poliedro l'oggetto considerato, usando un numero finito di vertici. Ovviamente andando a aumentare il numero di vertici l'approssimazione si avvicina sempre di più al caso reale e infatti uno dei primi obiettivi della tesi è capire quale sia un buon numero di vertici per approssimare i cilindri. Per questo il programma in C++ realizzato permette di scegliere il numero di vertici con cui approssimare il cilindro e li distribuisce sullo spigolo tondo di quest'ultimo.

Quindi, in sintesi, si vogliono studiare cilindri e dischetti tagliati approssimati con un numero finito di vertici e con delle patch attrattive posizionate in modo da simulare un filamento di DNA. Per maggiore chiarezza si presentano ora i due oggetti considerati e si discute in seguito come verificare l'intersezione o meno di questi oggetti.

4.1.1 Cilindro

Il primo modello considerato è quello del cilindro a corpo duro approssimato a poliedro come mostrato in figura 4.1. Il cilindro approssimato avrà diametro D e altezza h e, siccome i vertici vengono posizionati solamente sulla parte tonda, il poliedro avrà volume sempre inferiore a quello del cilindro. Più il numero di vertici aumenta più il volume si avvicinerà al volume originale. Nel caso della simulazione Monte Carlo di questi oggetti andrà verificata la non sovrapposizione (con gli algoritmi nel paragrafo 4.7) ma non andrà calcolata l'energia del sistema.

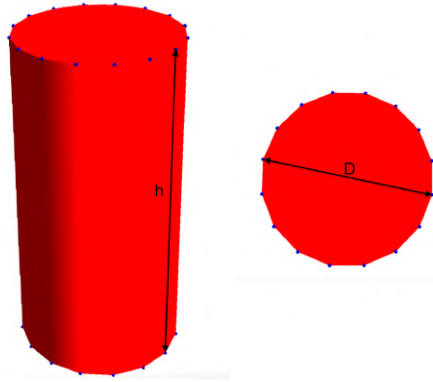


Figura 4.1. Vista laterale e vista superiore di un cilindro approssimato a poliedro con $V = 32$ vertici, con $D = 2$ e $h = 4$. Per maggiore chiarezza sono stati inseriti i vertici in blu, immagine realizzata utilizzando il programma *Molgl*.

4.1.2 Cilindro con 2 e 4 patch

In questo caso le particelle sono sempre dei cilindri approssimati a poliedri duri ma vengono inserite una o due patch per ogni base del cilindro. Nel caso con una singola patch per base, come mostrato in figura 4.2(a), questa viene inserita al centro della parte circolare del cilindro, quindi se questo è centrato nell'origine le patch avranno posizioni: $p_1 = (0, 0, +h)$ e $p_2 = (0, 0, -h)$.

Mentre nel caso di due patch per base queste vengono inserite a una distanza k , rispetto al centro della parte circolare. La scelta di k verrà spiegata nel dettaglio nei prossimi paragrafi, ma, sinteticamente, è dovuta al fatto che mettendole così distanziate si evitano interazioni multiple tra patch. La struttura a due patch per base è mostrata in figura 4.2(b).

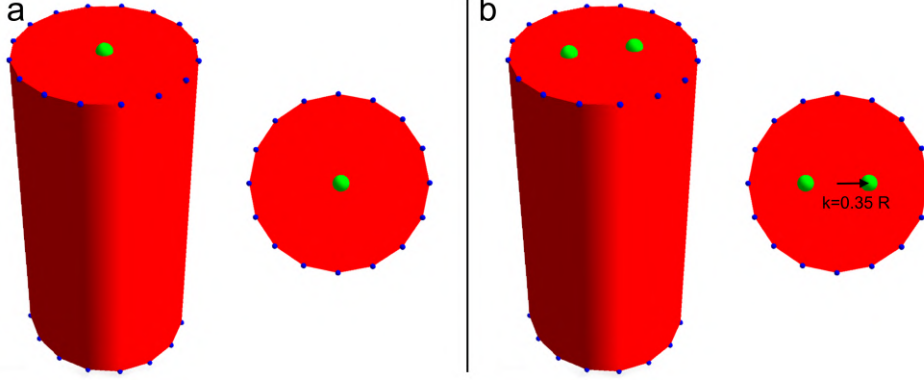


Figura 4.2. **a** Vista laterale e vista superiore di un cilindro approssimato a poliedro con $V = 32$ vertici, con $D = 2$ e $h = 4$ con una patch per base. **b** Caso analogo ma con 2 patch per base. Immagine realizzata utilizzando il programma *Molgl*.

Le patch attrattive interagiscono con un potenziale *square-well* u_{SW} , definito secondo la formula:

$$u_{SW} = \begin{cases} -u_0 & r \leq \delta_p \\ 0 & r > \delta_p \end{cases} \quad (4.1)$$

Dove r è la distanza tra le patch, u_0 è la profondità della buca di potenziale e δ_p è la distanza di interazione tra le patch, come mostrato in figura 4.3. Per convenienza, siccome T e u_0 nelle varie simulazioni compaiono sempre in rapporto, si definisce la temperatura ridotta adimensionale: $T^* = k_B T / U_0$.

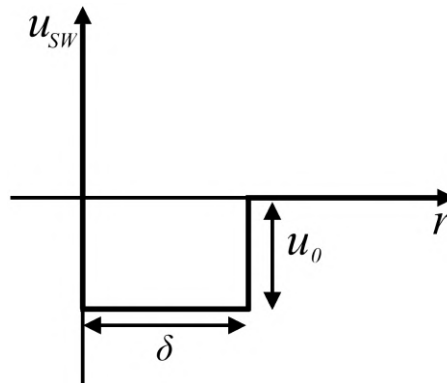


Figura 4.3. Grafico della buca di potenziale U_{SW} utilizzato sulle patch attrattive.

4.1.3 Semidisco con 4 patch

Le singole basi del DNA vengono approssimate con dei dischetti tagliati longitudinalmente con 4 patch totali, due sulla parte tagliata che saranno le patch di *pairing* e due sulle basi opposte che saranno le patch di *stacking*. Questa struttura è mostrata in figura 4.4.

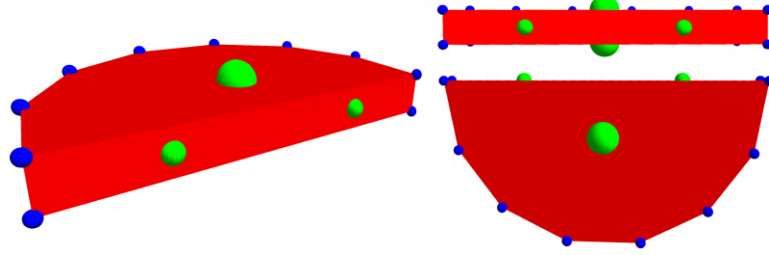


Figura 4.4. Vista laterale e superiore di un disco tagliato longitudinalmente approssimato a poliedro con inserite quattro patch. Il poliedro è stato approssimato utilizzando $V = 16$ vertici. Immagine realizzata utilizzando il programma *Molgl*.

La scelta delle dimensioni delle patch verrà giustificata nell'ottimizzazione della *persistence length* e della lunghezza finale media del filamento costruito. La posizione delle patch di *stacking* è giustificata anche dal non voler doppie interazioni e dal voler evitare costruzioni a Y (come mostrato in figura 4.5) causate da patch troppo distanti dal centro del cerchio di base.

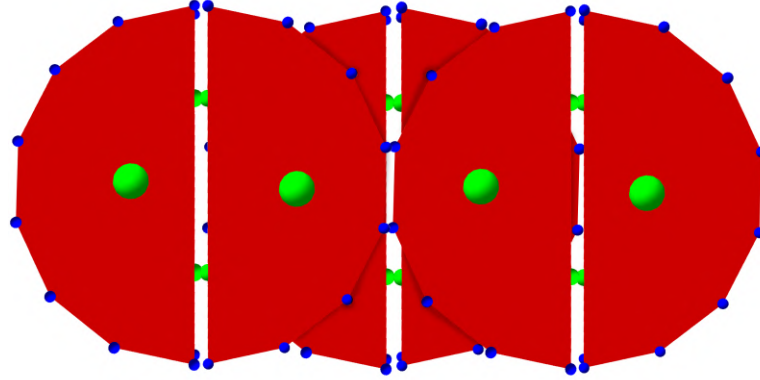


Figura 4.5. Immagine che rappresenta lo sdoppiamento del filamento a formare una Y, la patch qui è stata inserita a metà del raggio del cilindro e quindi permette questo fenomeno. Se la patch viene avvicinata al centro del cilindro è possibile evitare il fenomeno. Immagine realizzata utilizzando il programma *Molgl*.

Se si considera l'origine come il centro della faccia tagliata, il disco con altezza parallela all'asse \hat{z} e la faccia tagliata parallela al piano $\hat{x}\hat{z}$, le patch avranno posizione relativa al raggio r e altezza h pari a :

$$p_{stacking}^{1,2} = \begin{cases} (0, k \cdot r, h/2) \\ (0, k \cdot r, -h/2) \end{cases} \quad p_{pairing}^{1,2} = \begin{cases} (r/2, 0, 0) \\ (-r/2, 0, 0) \end{cases} \quad (4.2)$$

Dove k è la posizione della patch di *stacking* in relazione al raggio del disco, se $k = 0$ la patch sarà al centro del disco, mentre se $k = 1$ il centro della patch ricadrà sulla circonferenza. La figura 4.4 riporta un valore di $k = 0.35$, nel paragrafo 4.12.5 verrà discusso quale possa essere un valore ottimale di k e quali parametri della simulazione siano modificati da questa scelta. L'obiettivo della simulazione è quella di portare il sistema alla costruzione di filamenti di DNA, come mostrato in figura 4.6, quindi sarà necessario imporre alcune condizioni di esclusività sulle patch, discusse nel paragrafo A.1.

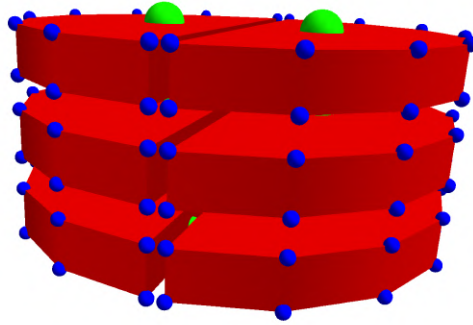


Figura 4.6. Filamento costruito impilando le basi formate da semidischi. In blu vengono rappresentati i vertici del poliedro che approssimano il semidisco, in questo caso sedici per oggetto. Immagine realizzata utilizzando il programma *Molgl*.

4.1.4 Semidisco con 4 patch e rotazione del filamento

La struttura presentata nello scorso paragrafo presenta un difetto. A differenza del filamento di DNA, la costruzione tramite semidischi aggiunge una anisotropia. Ad esempio se si guarda alla *persistence length*, che misura la flessibilità del filamento, si nota che questa varierà in base alla direzione in cui avviene la piega. Questo perché i tagli tra i vari semidischi sono tutti paralleli fra loro, questo può anche comportare che si crei un angolo tra due semidischi che formano un disco e che questo angolo si propaghi anche agli altri piani, a differenza del DNA dove i nucleotidi ruotano durante i legami di *stacking* dando così al filamento la classica struttura a doppia elica che previene questo fenomeno.

Come discusso nel paragrafo 2.3, il filamento di DNA compie una rotazione completa della doppia elica ogni 10 basi circa. È possibile costruire un modello che si avvicini ulteriormente alla struttura del DNA senza andare a complicare o aggiungere strutture geometriche, l'idea è quella di ruotare rispetto il centro le due patch di *stacking* in direzioni opposte di un angolo pari a $\pi/10$, come mostrato in figura 4.7. Le due patch avranno quindi un angolo tra loro di $\pi/5$ così da avere un filamento che ruoti su se stesso e compia una rotazione completa ogni 10 dischi, analogamente alla molecola di DNA.

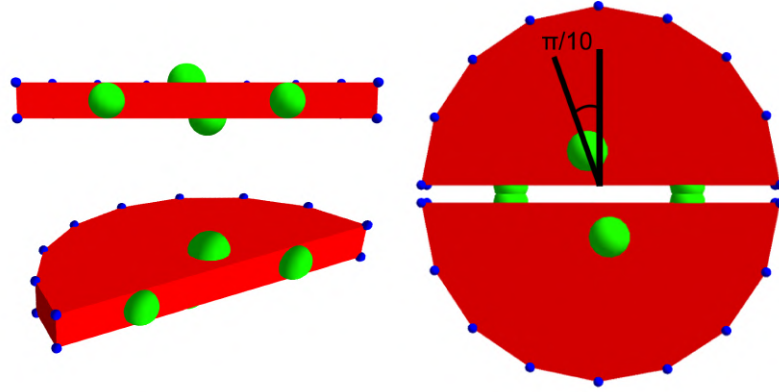


Figura 4.7. Filamento costruito inserendo le basi formate da semidischi, si inserisce tra le patch una rotazione di $\pi/5$ così da avere una rotazione completa ogni 10 basi. Immagine realizzata utilizzando il programma *Molgl*.

Questa nuova struttura, mostrata nella figura 4.8, non presenta direzioni preferenziali nella flessibilità e, nel caso si utilizzi per auto-assemblati, continua a funzionare come la precedente. Si decide quindi di utilizzare questo modello per le future approssimazioni visto che risolve il problema sulla flessibilità e non appesantisce le simulazioni in termini di tempi computazionali.

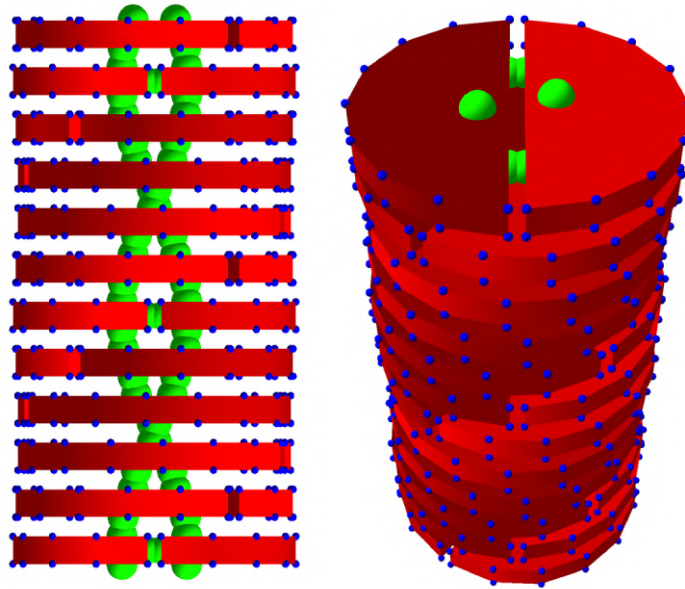


Figura 4.8. Filamento costruito inserendo le basi formate da semidischi, si inserisce tra le patch una rotazione di $\pi/5$ così da avere una rotazione completa ogni 10 basi. Immagine realizzata utilizzando il programma *Molgl*.

4.2 Efficienza delle mosse

Per ottimizzare l'efficienza delle simulazioni è necessario scegliere dei valori ottimali per i valori massimi di Δx , $\Delta\theta$ e ΔL che sono rispettivamente lo spostamento massimo di una particella, la rotazione massima e la variazione massima di un lato della box della scatola. Un valore troppo alto potrebbe creare configurazioni difficilmente accettabili per intersezioni tra particelle o alte energie, ma se troppo basso si rischia di far aumentare il tempo computazionale inutilmente [57]. Non c'è una probabilità fissa da ottenere ma dipende dal sistema considerato, in generale una buona regola è tenere una probabilità di accettazione compresa tra il 40% e il 60%. Per questo si è deciso di utilizzare il primo 15% dei passi per ottimizzare i valori di massima variazione, ovvero se l'accettazione è al di sotto del 40% si aumenta del 20% il valore massimo, altrimenti se l'accettazione è al di sopra del 60% si diminuisce del 20%.

Questo approccio è utile per ottimizzare i tempi ma presenta un difetto, la matrice α non è più simmetrica $\alpha(o \rightarrow n) \neq \alpha(n \rightarrow o)$ quindi non vale più la condizione di bilancio dettagliato. Di conseguenza è una routine utilizzabile solo nella parte iniziale della simulazione, e una volta fissati i valori, questi non possono più essere modificati.

4.3 Periodic boundary conditions

Attraverso le simulazioni molecolari si cerca di raccogliere informazioni sulle proprietà di un campione macroscopico. Tuttavia, la maggior parte delle simulazioni può gestire solo da poche centinaia a poche migliaia di particelle. Per eliminare l'effetto del sistema di dimensioni limitate e dell'effetto della superficie, e imitare la massa infinita che circonda il sistema di N particelle, si utilizzano nelle simulazioni le condizioni periodiche al contorno. Il problema risulta rilevante perché, se si prende per esempio un box sferico, si nota che il rapporto tra le particelle di superficie e le particelle interne ha un andamento come $N^{-1/3}$. Quindi se si prendono $N \approx 100$ particelle, il 22% di queste saranno particelle di superficie.

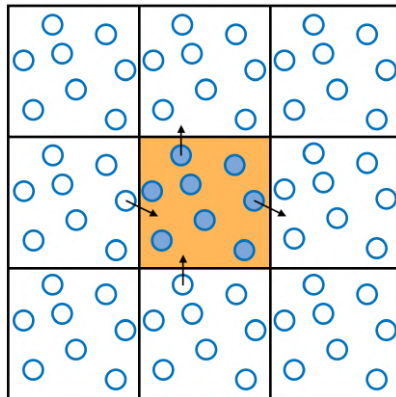


Figura 4.9. Schema bidimensionale del funzionamento delle PBC, il sistema di particelle viene ripetuto in tutte le celle vicine e se una particella esce dal box ricompare dal lato opposto.

Le condizioni periodiche al contorno sono mostrate nella Figura 4.9. Le particelle hanno immagini speculari in tutte le altre scatole. Una particella ora interagisce con tutte le altre particelle nel sistema infinito piuttosto che con un sistema finito. Se una particella esce dalla scatola verrà fatta rientrare dal lato opposto di essa. Per una data particella, viene presa in considerazione solo l'interazione con l'immagine periodica più vicina di qualsiasi altra particella, questa approssimazione viene detta convenzione di immagine minima ed è un'approssimazione ragionevole nel caso in cui si trattano interazioni a corto raggio.

La posizione della particella interagente andrà scalata, su uno o più assi, di una quantità pari alla dimensione della scatola. Di conseguenza la distanza tra una particella e l'immagine di una seconda particella più vicina alla prima viene calcolata come:

$$\begin{cases} \Delta r_x^{ij} = r_x^j - r_x^i - \left((int) \frac{r_x^j - r_x^i}{L_x/2} \right) * L_x \\ \Delta r_y^{ij} = r_y^j - r_y^i - \left((int) \frac{r_y^j - r_y^i}{L_y/2} \right) * L_y \\ \Delta r_z^{ij} = r_z^j - r_z^i - \left((int) \frac{r_z^j - r_z^i}{L_z/2} \right) * L_z \end{cases}$$

Dove con (int) si intende il casting a intero della quantità e la formula si riferisce al caso dove la scatola di simulazione è centrata nell'origine e quindi ha dimensione $[-L/2, L/2]$, nel caso la scatola fosse non centrata nell'origine $[0, L]$ andrà rimosso il fattore $1/2$ a denominatore. È possibile scegliere diverse forme per la scatola, tuttavia la forma scelta deve poter riempire tutto lo spazio con delle operazioni di traslazione, di conseguenza nel caso tridimensionale solamente alcune forme sono possibili, queste sono il cubo, il prisma esagonale, il dodecaedro rombico e l'ottaedro troncato (figura 4.10).

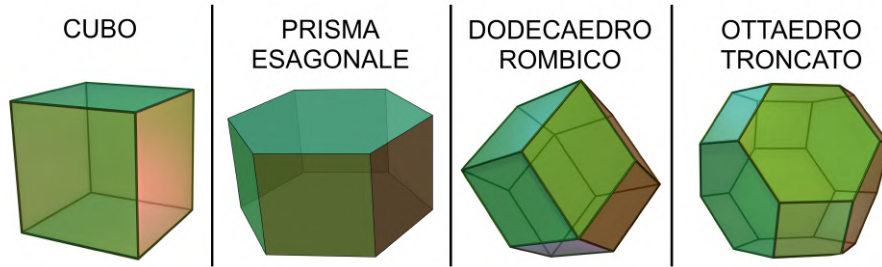


Figura 4.10. Poliedri tridimensionali che permettono il campionamento di tutto lo spazio attraverso delle operazioni di traslazione.

Se il potenziale di interazione presenta un raggio di cutoff r_C che taglia l'interazione a una certa distanza, le PBC richiedono che il lato della scatola di simulazione sia almeno il doppio di r_C , così che, nella sfera formata da r_C non sia presente una copia della particella su cui si sta calcolando l'interazione.

4.4 Random Number Generator RNG

La generazione di numeri casuali è un elemento fondamentale nell'ambito informatico, si utilizza in molte aree, tra cui la robotica, i videogiochi e le simulazioni. Le prestazioni di un generatore di numeri casuali sono definite dalla velocità con cui opera e dalla qualità dello schema di generazione di numeri casuali, qualità che è definita sia in ambito statistico ma anche, per esempio, in ambito di sicurezza ovvero quanto quanto sia prevedibile la sequenza una volta che ho già generato k numeri casuali. La prima volta che si utilizzarono i numeri generati casualmente per la risoluzione di problemi risale al 1949, ad opera di Nicholas Metropolis e Stanislaw Ulam, proprio per la realizzazione del metodo Metropolis [34], discusso nei precedenti capitoli. A seguire ci sono stati diversi sviluppi nell'area ma anche diversi passi falsi come il tristemente noto generatore *RANDU* [24].

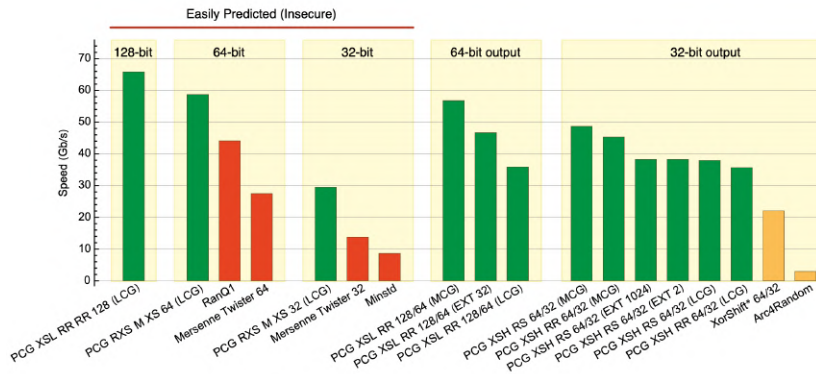


Figura 4.11. Benchmark delle performance dei generatori PCG contro i generatori maggiormente usati. Un numero più alto di Gb/s indica performance migliori. I generatori PCG sono tutti più veloci nelle rispettive classi e inoltre i generatori RanQ1, Mersenne Twister e Minstd falliscono i test statistici. Immagine tratta da pcg-random.org[40].

Nell'interesse della tesi la parte di crittografia, e quindi prevedibilità, dei numeri casuali viene accantonata visto che nelle simulazioni MC sono necessari dei generatori statisticamente forti e veloci nel caso di simulazioni con molte molecole. Si procede quindi a utilizzare gli algoritmi appartenenti alla famiglia PCG [40]. I vantaggi degli RNG della famiglia PCG sono la velocità, come mostrato in figura (4.11) e il rispetto dei test statistici. Nel caso di *pcg32* questo offre dei periodi di 2^{64} , le caratteristiche principali della famiglia di generatori di numeri casuali PCG sono mostrate nella tabella (4.12), insieme al miglior compilatore per ottimizzare i tempi di esecuzione.

Per confrontare la velocità e verificare la stabilità dell'algoritmo *pcg32* si effettuano due simulazioni NPT di cilindri duri di *aspect-ratio* $X_0 = 2$, nella prima si utilizza *pcg32* mentre nella seconda l'algoritmo *Mersenne Twister 64*. Entrambe le simulazioni equilibrano alla stessa frazione di volume, tuttavia la simulazione *pcg32* risulta il 3.5% più veloce. Considerando che nelle simulazioni Monte Carlo il tempo computazione è occupato, nella maggior parte, dagli algoritmi di intersezione tra le particelle, un miglioramento del 3.5% cambiando solo il metodo di generazione dei numeri casuali è un buon risultato. Si decide quindi di utilizzare l'algoritmo *pcg32*.

Method	Period	State (Bits)	Output (Bits)	Speed (ns/rand)	Speed (Gb/s)	Best Compiler
PCG XSL RR RR 128 (LCG)	2^{128}	128	128	1.81	65.86	g++ -O2
PCG RXS M XS 64 (LCG)	2^{64}	64	64	1.01	58.77	g++ -O2
RanQ1	$2^{64} - 1$	64	64	1.35	44.09	g++ -Os
Mersenne Twister 64	$2^{19937} - 1$	20032	64	2.17	27.48	g++ -O2
PCG RXS M XS 32 (LCG)	2^{32}	32	32	1.01	29.49	g++ -Os
Mersenne Twister 32	$2^{19937} - 1$	20032	32	2.16	13.79	g++ -O2
Minstd	$2^{31} - 2$	64	31	3.36	8.59	g++ -O2
PCG XSL RR 128/64 (MCG)	2^{126}	128	64	1.05	56.79	clang++ -O2
PCG XSL RR 128/64 (EXT 32)	2^{2174}	2176	64	1.29	46.36	clang++ -Os
PCG XSL RR 128/64 (LCG)	2^{128}	128	64	1.70	35.92	g++ -Os
PCG XSH RS 64/32 (MCG)	2^{62}	64	32	0.61	48.72	g++ -O2
PCG XSH RR 64/32 (MCG)	2^{62}	64	32	0.66	45.31	g++ -O2
PCG XSH RS 64/32 (EXT 1024)	2^{32830}	32832	32	0.78	38.29	g++ -O2
PCG XSH RS 64/32 (EXT 2)	2^{126}	128	32	0.78	38.29	g++ -O2
PCG XSH RS 64/32 (LCG)	2^{64}	64	32	0.79	37.95	g++ -O2
PCG XSH RR 64/32 (LCG)	2^{64}	64	32	0.84	35.67	g++ -O2
XorShift* 64/32	$2^{64} - 1$	64	32	1.35	22.05	g++ -Os
Arc4Random	$2^{1699} \sim$	2064	32	10.29	2.90	g++ -O2

Figura 4.12. Benchmark delle performance dei generatori PCG contro i generatori maggiormente usati. Immagine tratta da pcg-random.org[40].

4.5 Generazione delle configurazioni iniziali

Quando si effettua una simulazione ci sono diversi metodi per la generazione della configurazione iniziale, la scelta ottimale dipende da quale sistema si cerchi di simulare. Ad esempio si potrebbero generare N posizioni random \mathbf{r}_i nella scatola di simulazione, tuttavia questo metodo soffre del problema che potrebbero essere generate due posizioni molto vicine così da far risultare valori di energia molto alti. Una correzione possibile è di rifiutare tutte le mosse che portano a un salto dell'energia troppo grande o, alternativamente, si potrebbe definire una distanza minima tra la particella generata e tutte le altre particelle.

Questo permette di generare configurazioni disordinate, tuttavia non è ottimale a alte densità, siccome è sempre più difficile inserire la successiva particella e il numero di check da effettuare aumenta, aumentando, conseguentemente, il tempo computazionale. Un esempio di configurazione iniziale casuale è mostrato in figura 4.13 (a) per dei cilindri con 4 patch attrattive.

4.5.1 Configurazione iniziale cristallina

Se si cerca di ottenere configurazioni iniziali ad alte densità, l'alternativa è quella di generare le particelle su di un reticolo predefinito che può essere un reticolo cubico, cubico-centrato o anche cubico a facce centrate. Inoltre un altro vantaggio è che partire da configurazioni ordinate ad alta densità, permette di ottenere più velocemente fasi ordinate a più bassa densità. Un esempio di configurazione iniziale generata su di un reticolo cubico è mostrato in figura 4.13 (b). Va anche menzionata la possibilità di generare fasi liquido cristalline, come ad esempio la fase colonnare, per verificare che sia una fase stabile del sistema, questa in particolare è stata anche

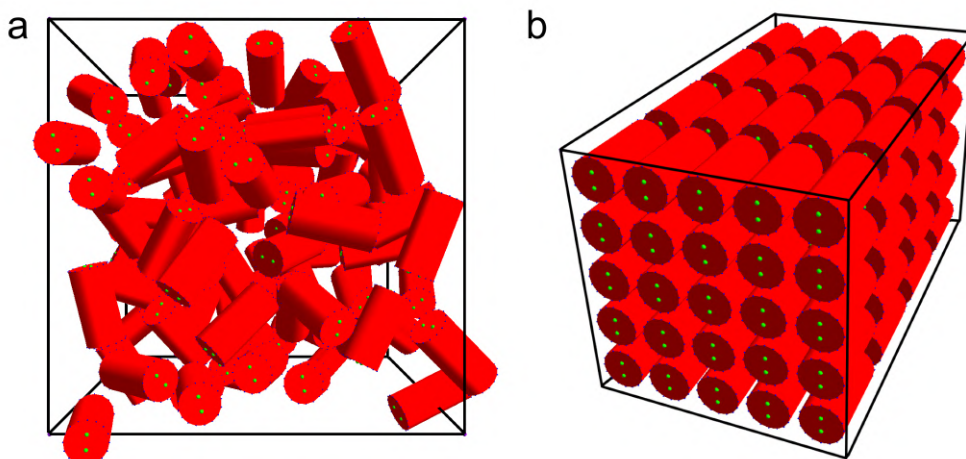


Figura 4.13. (a) Configurazione iniziale generata casualmente con delle molecole cilindriche con quattro patch. (b) Configurazione iniziale cristallina con cilindri di *aspect-ratio* 2. Immagine realizzata utilizzando il programma *Molgl*.

implementata nel codice che simula semidischi che verrà discussa in seguito. Il codice permette anche di partire da una configurazione fornita.

4.6 Impostazioni e ottimizzazioni del codice

Il codice scritto fornisce la possibilità di cambiare diverse impostazioni, in prima battuta è possibile scegliere la configurazione di partenza della simulazione di semidischi tra isotropa, cristallina e colonnare. Inoltre il codice permette di partire da configurazioni fornite così da poter realizzare un sistema di backup nel caso in cui il programma venga interrotto, salvando la configurazione raggiunta ogni un numero di passi Monte Carlo predefinito. È possibile decidere il comportamento delle patch, ovvero se queste devono mantenere i legami iniziali e se possono realizzare al massimo un legame. Si può agire anche sulla selettività delle patch permettendo solo interazioni di patch dello stesso tipo, *pairing* con *pairing* e *stacking* con *stacking*. Per maggiori informazioni riguardo alle impostazioni disponibili sulle patch si rimanda al paragrafo (A.1) in appendice.

Il codice viene ottimizzato utilizzando le *linked cell list*, il cui funzionamento è descritto in appendice (A.2). Vengono realizzate due liste separate, una per i semidischi e una per le singole patch attrattive.

4.7 Collisioni tra poliedri - SAT

Il rilevamento delle collisioni è un argomento molto ampio, rilevante per simulazioni molecolari e per altre applicazioni come la navigazione e la robotica. Questo paragrafo è stato scritto prendendo spunto dai libri [14] e [41]. La natura degli algoritmi di rilevamento delle collisioni dipende dai tipi di oggetti coinvolti e dalle informazioni necessarie al programma, ad esempio si può essere interessati a trovare solamente

se c'è una collisione o sapere anche in che zona questa avvenga. Si classificano gli algoritmi di rilevamento delle collisioni nelle seguenti ampie categorie:

- Oggetti stazionari, entrambi gli oggetti non si muovono. Qui si dividono in due ulteriori categorie, i programmi che testano solamente se ci sono intersezioni e quelli che rilevano anche le posizioni di tali intersezioni.
- Oggetti in movimento, uno o due oggetti sono in movimento, se entrambi sono in movimento è possibile scalare la velocità del secondo oggetto. Anche qui c'è la divisione tra programmi interessati solo a capire se c'è intersezione e programmi che trovano l'intersezione.

Ai fini della tesi verrà approfondito solo il rilevamento di intersezioni per oggetti statici, che verrà poi applicato nei vari programmi che discuteremo in seguito. Gli algoritmi di intersezione per oggetti generici possono essere estremamente complicati. Per questo motivo, si limita lo studio solamente agli oggetti convessi. Se S è un insieme di punti che rappresentano l'oggetto, l'insieme si dice convesso ogni volta che il segmento di linea che collega due punti dell'oggetto, X e Y , è a sua volta contenuto nell'insieme, indipendentemente dai due punti scelti. Cioè, se $X \in S$ e $Y \in S$, allora $(1 - t)X + tY \in S$ per ogni $t \in [0, 1]$. La figura (4.14) mostra due oggetti planari, uno convesso e uno non convesso.

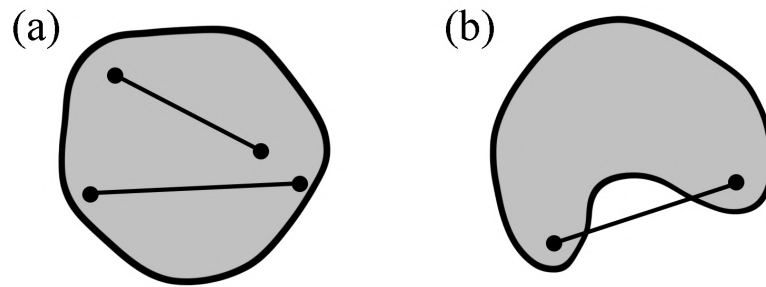


Figura 4.14. In figura (a) un oggetto convesso, la linea che collega due qualsiasi punti all'interno dell'oggetto è tutta contenuta all'interno. In figura (b) un oggetto non convesso, i due punti mostrati sono collegati da una linea che esce dall'oggetto.

Per gli oggetti che non sono convessi, un approccio tipico al calcolo dell'intersezione o della distanza consiste nel scomporre l'oggetto in un'unione di sotto-oggetti convessi (non necessariamente disgiunti) e applicare il programma per oggetti convessi a coppie tra due sotto-oggetti di oggetti differenti.

4.7.1 Separating Axis Theorem

Questa sezione descrive il Separating Axis Theorem (SAT) in 2D [13] [14], un metodo per determinare se due oggetti convessi stazionari sono sovrapposti o meno. Vengono presi in considerazione due tipi di query geometriche. La prima è una query test-intersezione che indica semplicemente se esiste o meno un'intersezione per oggetti fissi o se si verificherà per oggetti in movimento. La seconda è una query trova-intersezioni che comporta il calcolo dell'insieme delle intersezioni per due oggetti fissi o per due oggetti in movimento al momento del primo contatto. Ai

fini della tesi si descrive il primo tipo di query per poligoni convessi in 2D.

In questa sezione si utilizza la seguente notazione. Siano C_j per $j = 0$ i poligoni convessi con vertici $\{V_i^{(j)}\}_{i=0}^{N_j-1}$ ordinati in senso antiorario. I bordi dei poligoni sono $\vec{e}_i^{(j)} = V_{i+1}^{(j)} - V_i^{(j)}$ per $0 \leq i < N_j$ e dove $V_{N_j}^{(j)} = V_0^{(j)}$. Mentre i vettori uscenti normali alle facce del poliedro sono $\vec{d}_i^{(j)} = \text{Perp}(\vec{e}_i^{(j)})$, dove $\text{Perp}(x, y) = (-y, x)$.

4.7.2 Proiezione su una linea

Il check da effettuare per verificare la non intersezione di due oggetti convessi è: se esiste una linea per la quale gli intervalli di proiezione dei due oggetti su quella linea non si intersecano, allora gli oggetti non si intersecano. Tale linea è chiamata linea di separazione o, più comunemente, asse di separazione (si veda la Figura (4.15)).

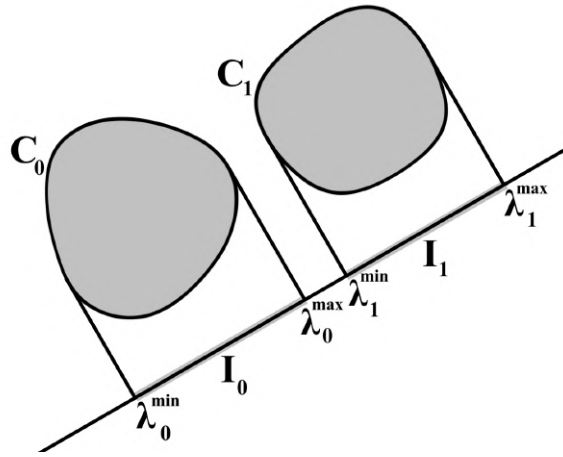


Figura 4.15. Proiezione su un asse di separazione di due oggetti convessi separati.

La traslazione di una linea di separazione è anche, a sua volta, una linea di separazione. Data una retta contenente l'origine e con direzione di lunghezza unitaria \vec{d} , la proiezione di un insieme convesso C sulla retta è l'intervallo:

$$I = [\lambda_{\min}(\vec{d}), \lambda_{\max}(\vec{d})] = [\min\{\vec{d} \cdot \vec{X} : \vec{X} \in C\}, \max\{\vec{d} \cdot \vec{X} : \vec{X} \in C\}] \quad (4.3)$$

dove eventualmente $\lambda_{\min}(\vec{d}) = -\infty$ o $\lambda_{\max}(\vec{d}) = +\infty$; questi casi sorgono quando l'insieme convesso è illimitato. Due insiemi convessi C_0 e C_1 sono separati se esiste una direzione \vec{d} tale che gli intervalli di proiezione I_0 e I_1 non si intersecano. In particolare non si intersecano quando:

$$\lambda_{\min}^{(0)}(\vec{d}) > \lambda_{\max}^{(1)}(\vec{d}) \quad \text{o} \quad \lambda_{\max}^{(0)}(\vec{d}) < \lambda_{\min}^{(1)}(\vec{d}) \quad (4.4)$$

Siccome la proiezione è invariante per rette parallele la scelta di \vec{d} rimane valida se si prende un vettore scalato $t\vec{d}$ o se si prende la direzione opposta $-\vec{d}$.

4.7.3 Poligoni convessi stazionari in 2D

Per i test di separazione su una coppia di poligoni convessi deve essere considerato solo un insieme finito S di vettori di direzione. Quel set, nel caso bidimensionale

contiene solo i vettori normali ai bordi dei poligoni. La Figura 4.16 (a) mostra due poligoni non intersecanti che sono separati lungo una direzione determinata dalla normale al bordo di un poligono. La figura 4.16 (b) mostra due poligoni che si intersecano; non ci sono direzioni di separazione.

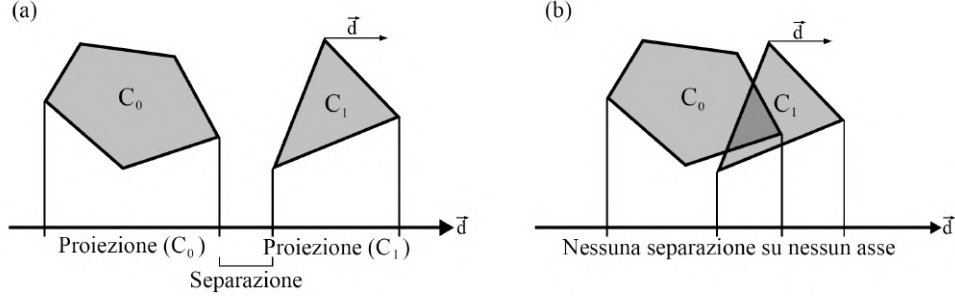


Figura 4.16. (a) Proiezione, su un asse di separazione, di due oggetti convessi separati. (b) Proiezione su di un asse di due oggetti compenetrati.

Per capire la motivazione per cui devono essere testate solo le normali agli spigoli si guarda il caso limite, ovvero due poligoni convessi che si toccano appena senza compenetrazione. La Figura 4.17 mostra le tre possibili configurazioni: contatto bordo-bordo, contatto vertice-bordo e contatto vertice-vertice. Le linee che separa i poligoni nell'immagine è perpendicolare alle linee di separazione che si verificherebbe per un oggetto traslato dall'altro di una distanza infinitesima.

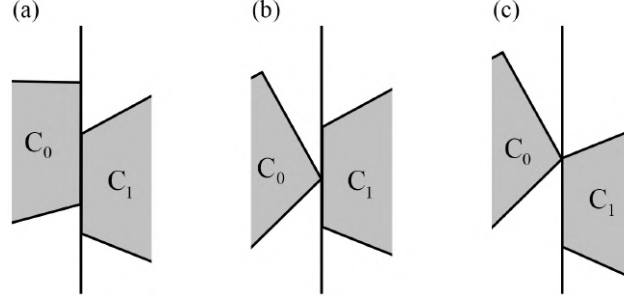


Figura 4.17. (a) Contatto bordo-bordo. (b) Contatto vertice-bordo. (c) Contatto vertice-vertice

La prova matematica che S contiene solo normali di spigolo si basa sulla visualizzazione che se \vec{d} è una direzione di separazione che non è normale a un bordo di nessuno dei due poligoni convessi allora ci deve essere una normale di uno spigolo che sia anche una direzione di separazione. Sia $\vec{d} = (\cos \theta, \sin \theta)$ una direzione di separazione non normale a un bordo. Per semplicità, si supponga che la proiezione di C_0 sulla linea di separazione sia a sinistra della proiezione di C_1 , come mostrato in figura 4.18 (a). Nel caso fosse a destra il discorso sarebbe analogo. Dal momento che \vec{d} non è una normale del bordo, solo un vertice V_0 di C_0 viene proiettato su $\lambda_{max}^{(0)}$ e, analogamente solo un vertice V_1 di C_1 viene proiettato su $\lambda_{min}^{(1)}$. Sia θ_0 l'angolo più grande minore di θ in modo che $\vec{d}_0 = (\cos \theta_0, \sin \theta_0)$ è una normale al bordo, ma $\vec{d}(\phi) = (\cos \phi, \sin \phi)$ non è una normale al bordo per $\phi \in (\theta_0, \theta]$. Allo stesso modo,

sia θ_1 l'angolo più piccolo maggiore di θ così che $\vec{d}_1 = (\cos \theta_1, \sin \theta_1)$ è una normale al bordo, ma $\vec{d}(\phi)$ non è una normale al bordo per $\phi \in [\theta_0, \theta_1)$. Per tutte le direzioni $\vec{d}(\phi)$ con $\phi \in (\theta_0, \theta_1)$, V_0 è l'unico vertice che viene proiettato su $\lambda_{min}^{(1)}$. La separazione tra gli intervalli è la funzione continua $f(\phi) = (\cos \phi, \sin \phi) \cdot (V_1 - V_0) = A \cos(\phi + \psi)$, dove A è l'ampiezza costante e ψ è l'angolo di fase costante. Inoltre, $f(\theta) > 0$ siccome \vec{d} è un asse di separazione. Le variabili in gioco, \vec{d}_0 , \vec{d}_1 e $f(\theta)$, sono mostrate in figura 4.18 (a).

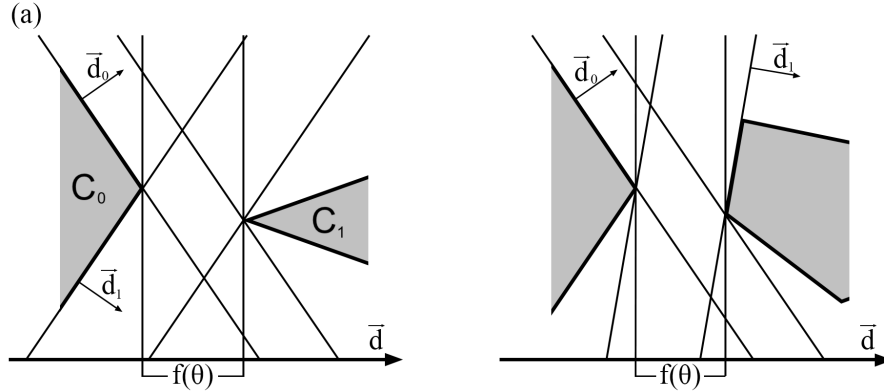


Figura 4.18. Le normali al bordo più vicine a una direzione di separazione non normale al bordo: (a) dallo stesso triangolo e (b) da triangoli diversi.

Se $f(\theta_0) > 0$, allora la normale al bordo \vec{d}_0 è una direzione di separazione. Se $f(\theta_1) > 0$, allora anche la normale al bordo \vec{d}_1 è una direzione di separazione. Supponiamo che $f(\theta_0) \leq 0$ e $f(\theta_1) \leq 0$. Poiché $f(\theta) > 0$, devono esistere due zeri di f su $[\theta_0, \theta_1]$, uno minore di θ e uno maggiore di θ . Gli zeri di f sono separati da π radianti. Questo forza $\theta_1 - \theta_0 \geq \pi$, in questo caso l'angolo tra le normali di bordo consecutive \vec{d}_0 e \vec{d}_1 è almeno π radianti. Ciò accade solo se l'angolo è esattamente π , i due spigoli che condividono V_0 sono paralleli a \vec{d} , e i due spigoli che condividono V_1 sono anch'essi paralleli a \vec{d} , questo è in contraddizione al fatto che gli angoli siano strettamente positivi. Pertanto si è dimostrato per assurdo che è impossibile siano entrambe vere $f(\theta_0) \leq 0$ e $f(\theta_1) \leq 0$.

In sintesi, se $f(\theta) > 0$, allora o $f(\theta_0) > 0$, in questo caso \vec{d}_0 è una normale di separazione a un bordo, oppure $f(\theta_1) > 0$, in questo caso \vec{d}_1 è una normale di separazione a un bordo. La Figura 4.18 (a) mostra il caso in cui entrambe le normali agli spigoli più vicine provengano dallo stesso oggetto. La Figura 4.18 (b) mostra il caso in cui le normali agli spigoli più vicine provengano da oggetti diversi.

Riassumendo, il test di separazione per oggetti 2D convessi consiste nella verifica di non sovrapposizione delle proiezioni dei vertici su tutte le normali dei lati di entrambi gli oggetti. Di conseguenza, se si considera il caso più semplice con due triangoli come mostrato in Figura (4.19), le proiezioni in totale da fare sono su 6 direzioni. Il vantaggio di questa tecnica risulta nel fatto che, trovata una direzione di separazione, gli oggetti saranno sicuramente separati e non è quindi necessario procedere con le proiezioni nelle direzioni rimanenti.

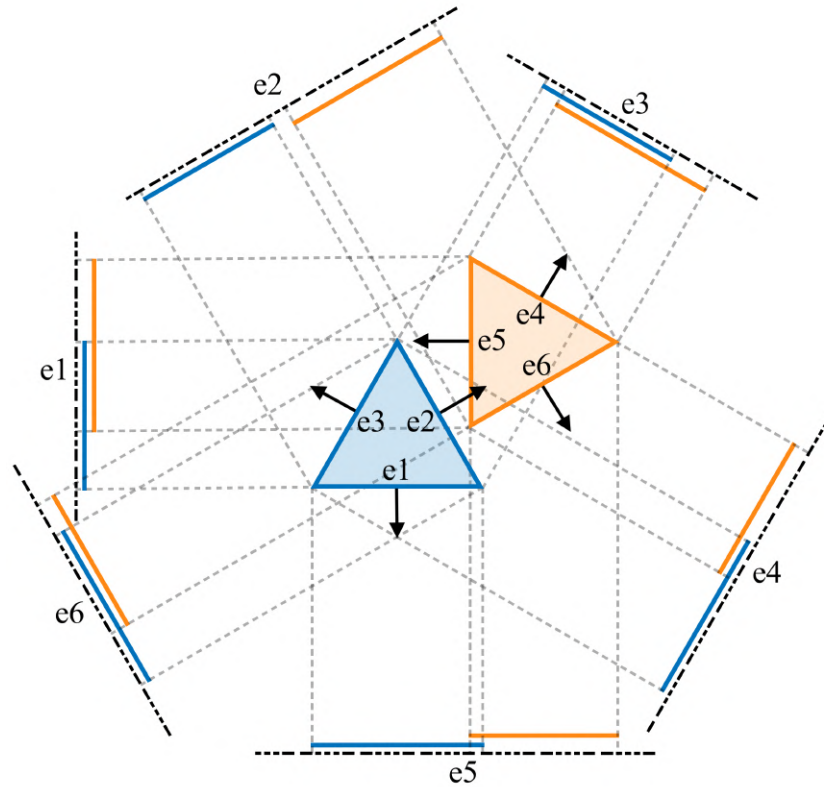


Figura 4.19. Applicazione del teorema della separazione degli assi su due triangoli equilateri, la separazione dei due oggetti è visibile solo nella proiezione sulla retta parallela al vettore $e2$.

Infatti, nell'esempio in Figura (4.19), una volta verificate le direzioni $e1$ e $e2$ l'algoritmo può uscire riportando la non intersezione senza testare le altre quattro direzioni rimanenti. Nel caso peggiore in termini di tempi computazionali il numero di assi da testare sarà pari alla somma del numero delle facce di entrambi gli oggetti, che nel caso di due triangoli sono sei. Se la separazione non viene trovata in queste sei direzioni allora i due oggetti saranno sovrapposti.

Si noti inoltre che gli assi paralleli non devono essere testati più volte. Ad esempio, se si considera il caso in cui si hanno due quadrati è necessario proiettare i vertici solo su due direzioni per quadrato, visto che le normali delle facce sono parallele a coppie sul singolo quadrato.

4.7.4 Poligoni convessi stazionari in 3D

Nel caso tridimensionale, a differenza di quello bidimensionale, avviene una distinzione tra spigoli e facce degli oggetti considerati. Di conseguenza è necessario modificare l'algoritmo per tener conto delle interazioni spigolo-spigolo e evitare falsi positivi di intersezione. Un esempio, del caso tridimensionale, dove usare solamente le normali alle facce non è sufficiente è mostrato in figura (4.20).

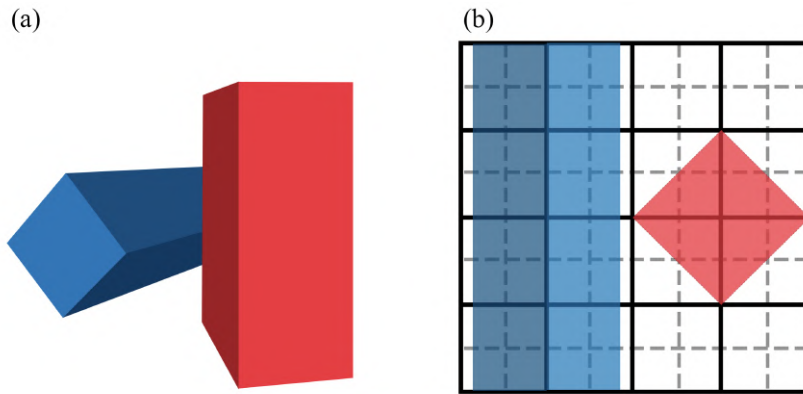


Figura 4.20. Caso in cui l'algoritmo che controlla solamente le normali alle facce rileva una sovrapposizione anche con oggetti separati, in questo caso due cuboidi in (a) vista laterale e (b) vista dall'alto.

I due cuboidi (in rosso e blu) non sono sovrapposti, come si nota dalla vista dall'alto, tuttavia utilizzando la versione del paragrafo precedente del separating axis theorem viene rilevata una sovrapposizione, quindi l'algoritmo produce un falso positivo rilevando una collisione quando non è presente. Questo processo è illustrato in Figura (4.21), dove si mostra che non è presente una direzione che mostra la separazione dei due cuboidi.

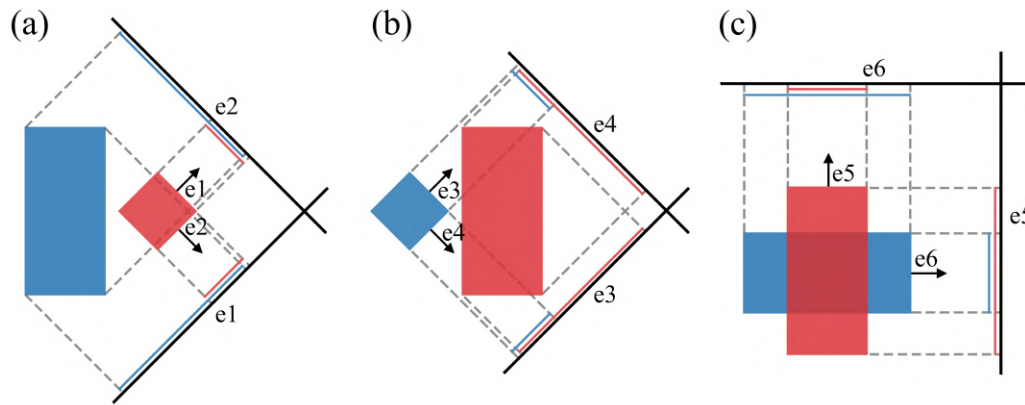


Figura 4.21. Il test del separating axis theorem solo sulle normali alle facce dei cuboidi non rileva la separazione dei due oggetti. (a) Vista dall'alto, (b) vista frontale e (c) vista laterale.

Per risolvere il problema è necessario proiettare su assi ulteriori così da trovare la direzione in cui i due oggetti siano separati. Le direzioni da controllare per rilevare le interazioni spigolo-spigolo sono tutte le direzioni perpendicolari a tutte le possibili coppie di spigoli. Così facendo verrebbe rilevata anche la separazione mostrata in Figura (4.20) risolvendo il problema del falso positivo.

Effettuando il prodotto vettoriale di uno spigolo dell'oggetto 1 con uno spigolo

dell'oggetto 2 verrà prodotto un asse ortogonale a entrambi che potrà essere uno degli assi di separazione. Questa implementazione assicurerà di rilevare tutte le collisioni tra poliedri convessi senza falsi positivi.

Tuttavia questa modifica all'algoritmo fa sì che il tempo computazionale per la verifica di un overlap scali con il cubo del numero di vertici, infatti se prima era necessario proiettare i poliedri solo in un numero di direzioni pari al numero di facce, ora a queste si aggiungono tutte le permutazioni dei prodotti vettoriali di due spigoli. Per questo risulta importante ottimizzare il codice sfruttando le simmetrie del poliedro considerato non andando a proiettare su assi paralleli tra loro, di questo si discuterà nel prossimo paragrafo.

4.7.5 Implementazione e ottimizzazione codice

L'ottimizzazione del codice si basa sullo sfruttare le simmetrie dell'oggetto di cui trovare le collisioni, come discusso nel paragrafo precedente, non è necessario proiettare su assi paralleli a assi su cui si è già proiettato.

Intersezione tra cuboidi

Si consideri ad esempio il caso in cui si voglia rilevare la collisione tra due cuboidi, si dovranno testare prima gli assi paralleli alle normali alle facce e poi gli assi derivanti dal prodotto vettoriale degli spigoli. Per prima cosa si considerino gli assi paralleli alle normali alle facce, in teoria qui andrebbero testati un numero di assi pari al numero delle facce quindi sei dal primo cuboide e sei dal secondo. Tuttavia sul singolo cuboide le normali alle facce saranno vettori paralleli con opposta direzione quindi di questi ne andranno considerate solo la metà.

Se si considera il prodotto vettoriale degli spigoli si nota, in prima battuta, che il prodotto vettoriale cambia solo direzione scambiando l'ordine dei due vettori considerati, di conseguenza sarà necessario solo testare tutti gli spigoli del primo cuboide con ogni spigolo del secondo e non il viceversa. Ulteriormente si nota che gli spigoli paralleli di ognuno dei due cuboidi non devono essere considerati se non una volta, perché una volta fatto il prodotto vettoriale l'asse risultante sarà sempre lo stesso. Se si considera un cuboide il numero di spigoli non paralleli tra loro è pari a tre, discorso analogo per il secondo cuboide. Si dovrà quindi effettuare il test su $3 \times 3 = 9$ assi, a differenza del caso in cui si effettua il test tutti gli spigoli in cui si avrebbero $8 \times 8 = 64$ assi. Nel caso di cuboidi si è passati quindi da effettuare un totale di 76 proiezioni su assi a doverne fare 15, ottenendo lo stesso risultato.

Intersezione tra cilindri

Analogamente, se si è interessati a verificare l'intersezione di due cilindri approssimati a poliedri (mostrati in figura 4.1), è possibile ottimizzare il numero di proiezioni su assi da effettuare. Si considera il caso in cui i cilindri siano approssimati con $V = 32$ vertici. Le facce della parte tonda non devono essere testate tutte nella proiezione ma solamente la metà, siccome a coppie queste saranno antiparallele.

Per lo stesso motivo va considerata solamente una normale per le due basi. Nel prodotto vettoriale degli spigoli vale la stessa osservazione che per i cuboidi, ovvero è necessario solo fare il prodotto vettoriale degli spigoli di un poliedro con tutti gli spigoli del secondo e non il viceversa. Ulteriormente, non è necessario testare tutti gli spigoli paralleli all'asse del cilindro della parte circolare nel prodotto vettoriale perché paralleli fra loro, permettendo una velocizzazione notevole nella simulazione. Per ultimo vanno considerati solo la metà dei spigoli della base perché paralleli a coppie e inoltre, sempre perché paralleli a coppie, va considerata solo una base.

Ricapitolando andranno testate 9 normali per poliedro invece che 18. Mentre, senza ottimizzazioni, andrebbero testati il prodotto vettoriale di 48 spigoli per poliedro e quindi un totale di 2304 rette su cui proiettare. Utilizzando le ottimizzazioni si dovranno considerare solamente 9 vertici per poliedro, portando così a fare 81 proiezioni su rette. In definitiva si passa a fare un totale di 2340 proiezioni a farne 99. Nel paragrafo 4.10 si confronteranno i tempi di esecuzione dei due programmi.

Intersezione tra dischetti tagliati

Questo caso è analogo a quello dei cilindri, si consideri il caso in cui si approssima il dischetto tagliato con $V = 18$ vertici. In questo caso vanno considerate tutte le facce del poliedro tranne una, visto che le basi sono antiparallele. Mentre gli spigoli da considerare nel prodotto vettoriale sono 10 invece che 27, passando a 100 proiezioni da 729.

Ottimizzazione proiezione su retta di poliedri

É possibile ottimizzare il processo di proiezione su retta di due poliedri, per prima cosa si cerca il vertice del primo poliedro con la proiezione più vicina alla proiezione del centro del secondo poliedro V_{max}^1 e, una volta che si proietta il secondo poliedro, ci si ferma se si trova una proiezione più vicina al centro del primo poliedro rispetto a V_{max}^1 . In questo caso non è necessario continuare a proiettare perché i poliedri saranno sicuramente non separati su questo asse, il processo è mostrato in figura 4.22. Nel caso in cui si stia proiettando su un asse di separazione l'algoritmo dovrà comunque proiettare tutti i vertici.

Si può ottimizzare la proiezione sfruttando la struttura a cilindro e a dischetto tagliato. Nel momento in cui si proiettano i vertici si procede in ordine su una base (circolare o semi-circolare a seconda del poliedro) e non appena si trova una proiezione che è più vicina al centro del secondo poliedro rispetto la proiezione del vertice precedente e successivo ci si ferma. Trovata la proiezione massima su una base V_{max}^1 non è necessario testare tutta la base superiore siccome, per geometria del sistema, la proiezione massima, se c'è, sarà il vertice corrispondente a V_{max}^1 . Così è possibile dimezzare il numero di vertici da proiettare. Si procederà poi analogamente nella proiezione del secondo poliedro.

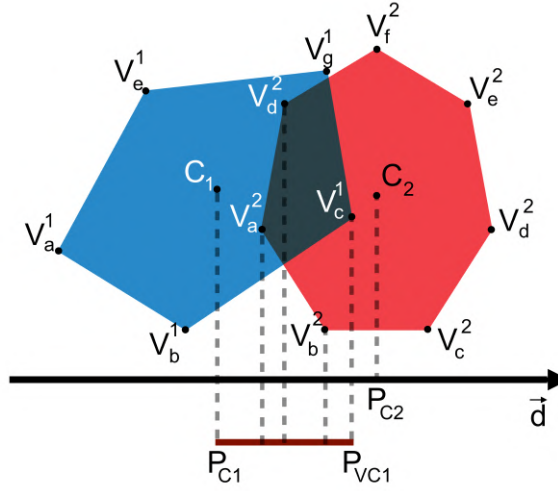


Figura 4.22. Trovata la proiezione del poliedro 1 più vicina al centro del poliedro 2 C_2 , in questo caso V_C^1 , si procede con la proiezione dei vertici del secondo poliedro. Il programma può uscire e riportare un'intersezione se trova una proiezione del vertice del poliedro 2 che cada nell'intervallo $(-\infty, P_{VC1})$. Nell'immagine cadono in questo intervallo i vertici V_a^2 , V_b^2 e V_d^2 .

4.8 Collisione tra poliedri - *Xenocollide*

Si presenta ora un secondo algoritmo per verificare l'intersezione di un gran numero di oggetti convessi, la scelta di utilizzare un secondo algoritmo deriva dalla velocità di quest'ultimo, infatti nel confronto fatto nel paragrafo 4.10 risulta evidente il vantaggio visto che scala linearmente con il numero di vertici. Questo algoritmo si chiama *Xenocollide*[22] e appartiene alla classe di algoritmi basati sulla tecnica *Minkowski Portal Refinement* (MPR). *Xenocollide* condivide molte similitudini con l'algoritmo di rilevamento di intersezioni GJK[19], anche se quest'ultimo risulta meno robusto. Si presentano ora la mappatura di supporto e le differenza di Minkowski.

4.8.1 Mappatura di supporto

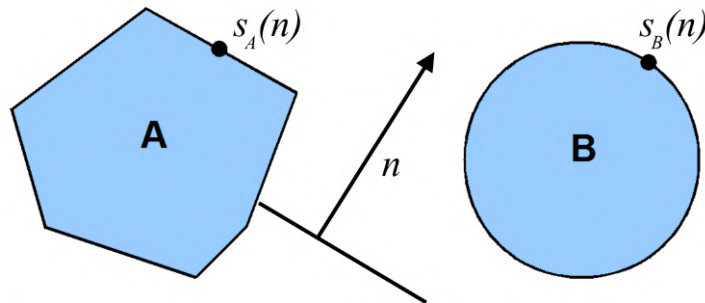


Figura 4.23. Rappresentazione della funzione mappatura di supporto per due oggetti bidimensionali. I punti $s_A(n)$ e $s_B(n)$ saranno i due punti di supporto selezionati.

La base degli algoritmi che funzionano con una vasta gamma di forme geometriche è quella di trovare un metodo generale per rappresentarle, in questo caso si utilizza la mappatura di supporto per rappresentare oggetti convessi. La mappatura di supporto è una funzione matematica che prende un vettore direzione in input e restituisce il punto, appartenente a un oggetto convesso, la cui proiezione si trova più avanti in quella direzione. Nel caso ci siano più punti a soddisfare tale richiesta allora se ne può scegliere uno qualsiasi a patto che la scelta sia consistente, ovvero utilizzando più volte la mappatura di supporto con gli stessi input l'output non cambi. Un esempio della funzione di mappatura di supporto è mostrato in figura 4.23 per una direzione n e due oggetti convessi. Il piano ortogonale alla direzione n è chiamato piano di supporto, tutti i punti degli oggetti che si trovano sul lato opposto del piano rispetto a n non vengono considerati nella mappatura.

Nel caso di una sfera posta nell'origine la mappatura di supporto può essere scritta come una funzione, ovvero $\mathbf{S}_{sfera}(\mathbf{n}) = r \mathbf{n}$. Dove r è il raggio della sfera. Questa funzione può essere trovata per diverse forme geometriche, nel caso di una box centrata nell'origine e con spigoli paralleli agli assi del sistema sarà:

$$\mathbf{S}_{box}(\mathbf{n}) = [r_x \text{sgn}(n_x), r_y \text{sgn}(n_y), r_z \text{sgn}(n_z)] \quad (4.5)$$

Questo nel caso in cui gli oggetti siano centrati nell'origine e allineati con gli assi del sistema di riferimento, la mappatura di supporto per un oggetto ruotato e traslato nello spazio è data da:

$$\mathbf{S}_{lab}(\mathbf{n}) = \mathbf{R} \mathbf{S}_{obj}(\mathbf{R}^{-1} \mathbf{n}) + \mathbf{T} \quad (4.6)$$

Dove con \mathbf{R} si indica la matrice di rotazione dell'oggetto mentre con \mathbf{T} si indica il vettore di traslazione per il cambio di sistema di riferimento.

4.8.2 Differenza di Minkowski

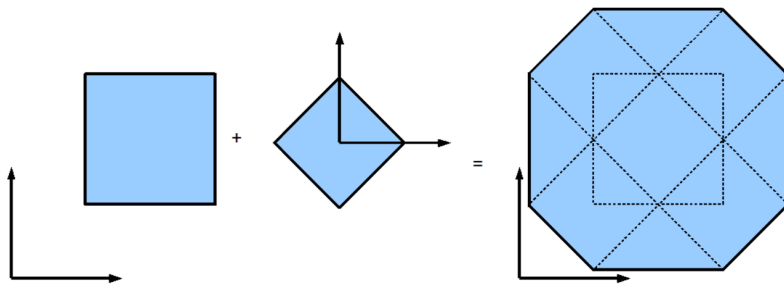


Figura 4.24. Rappresentazione della differenza di Minkowski per due oggetti bidimensionali. Immagine tratta da [19].

Se si sottraggono a tutti i punti di un oggetto convesso A tutti i punti di un secondo oggetto convesso B si ottiene la forma chiamata differenza di Minkowski, mostrata nel caso bidimensionale nella figura 4.24. Se A e B sono oggetti convessi allora anche la loro differenza sarà un oggetto convesso. Se i due oggetti si sovrappongono ci sarà allora almeno un punto contenuto sia in A che in B nel sistema di

riferimento del laboratorio. Quindi se la differenza di Minkowski $B - A$ contiene l'origine allora A e B si sovrappongono. Altrimenti se la differenza non contiene l'origine allora A e B sono separati.

Effettuare la differenza su tutti i punti di un oggetto è costoso in termini di tempi computazionali. È possibile calcolare la mappatura di supporto della differenza di Minkowski utilizzando l'equazione 4.7 e determinare l'intersezione di A e B controllando se l'origine è contenuta in S_{B-A} .

$$S_{B-A}(\mathbf{n}) = S_B(\mathbf{n}) - S_A(\mathbf{n}) \quad (4.7)$$

4.8.3 Rilevamento di collisioni - 2D *Minkowski Portal Refinement*

Una volta costruita la differenza di Minkowski bisogna verificare se l'origine sia contenuta in essa, l'algoritmo *XenoCollide* [54], serve a ottimizzare questa ricerca e per farlo utilizza il *Minkowski Portal Refinement* MPR. Si procede in questo paragrafo a mostrare il procedimento nel caso bidimensionale.

Definizione del portale

Si consideri il caso in cui la differenza di Minkowski, $B - A$, sia rappresentata da un ettagono e sia O l'origine di tale sistema come mostrato in figura 4.25 (a). L'obiettivo dell'algoritmo è determinare se l'origine sia contenuta in $B - A$, utilizzando solo la mappatura di supporto per rappresentare la superficie.

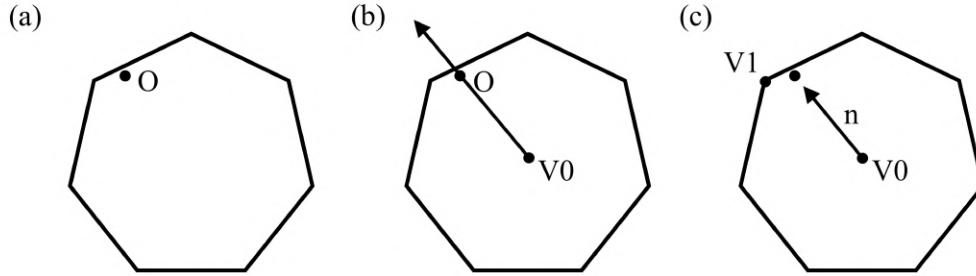


Figura 4.25. (a) Definizione di $B - A$ e dell'origine. (b) Semiretta che unisce $V0$ all'origine. (c) Utilizzando la funzione di supporto si individua il punto $V1$.

Inizialmente si sceglie un punto qualsiasi contenuto all'interno di $B - A$, per farlo basta sottrarre a un punto contenuto in A un punto contenuto in B . Siccome la velocità con cui l'algoritmo riporta una intersezione o meno dipende da questa scelta si scelgono i centri geometrici di A e B . Questo punto viene indicato con $V0$ nella figura 4.25 (b).

Si costruisce la normale che parte dal punto $V0$ e punta verso l'origine. Si quindi trova il punto di supporto $V1$ in tale direzione come mostrato in figura 4.25 (c).

Successivamente si costruisce il vettore perpendicolare al segmento che unisce il punto di supporto $V1$ a $V0$. Per scegliere il verso in cui il vettore punta si cerca

quello che giace sullo stesso lato dell'origine rispetto al segmento. Si utilizza quindi questo vettore per trovare il secondo punto di supporto della superficie $B - A$, come mostrato in figura 4.26 (a).

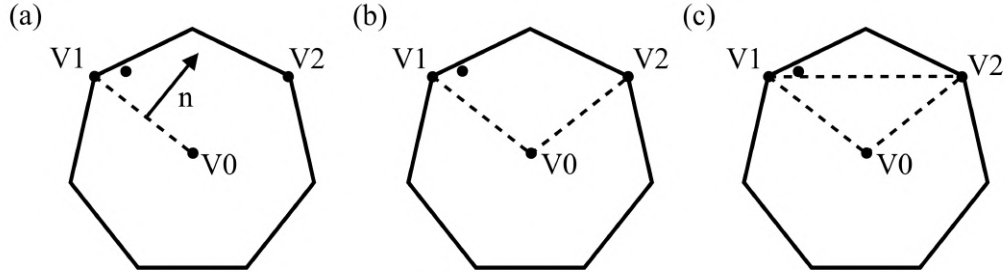


Figura 4.26. (a) Si definisce la normale n e si trova il punto $V2$ con la funzione di supporto. (b) L'origine è contenuta nell'angolo $V1\widehat{V0}V2$. (c) L'origine è esterna al triangolo formato da $V0$, $V1$ e $V2$. $V1$ e $V2$ formano il portale.

I tre punti trovati definiscono l'angolo $V1\widehat{V0}V2$, l'origine sarà all'interno di tale angolo come si nota in figura 4.26 (b).

Come mostrato in figura 4.26 (c), si unisce i due punti di supporto, $V1$ e $V2$, con un segmento. Questo segmento è chiamato portale siccome la retta che unisce $V0$ all'origine dovrà per costruzione attraversarlo.

Perfezionamento del portale

Se l'origine giace sullo stesso lato del portale del punto $V0$ allora sarà contenuta nel triangolo tratteggiato formato dagli angoli $V0$, $V1$ e $V2$ mostrato in figura 4.26 (c). In questo caso l'algoritmo riporta una intersezione tra i due oggetti A e B . Nell'esempio qui mostrato l'origine è esterna al portale quindi l'algoritmo continua.

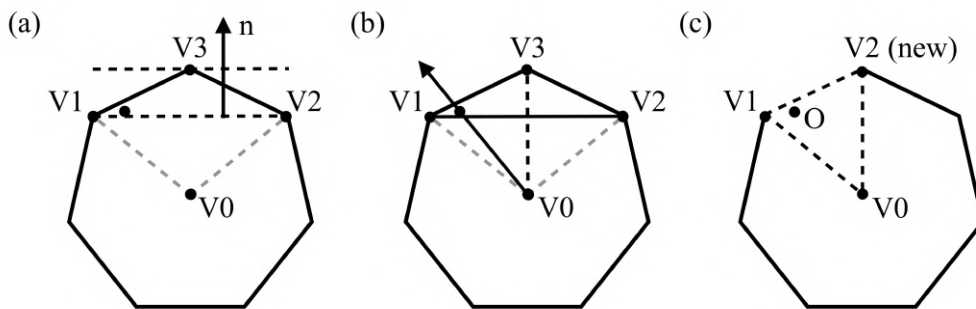


Figura 4.27. (a) Si utilizza la funzione di supporto nella direzione n individuando il punto $V3$. (b) Si mostra il triangolo formato da i punti $V1$, $V2$ e $V3$. La semiretta da $V0$ e l'origine attraversa la base del triangolo e esce da uno degli altri due lati. (c) Il lato $V1V2$ del triangolo diviene il nuovo portale e l'algoritmo riporta una intersezione.

Si costruisce il vettore perpendicolare al portale che punta verso il lato opposto a $V0$. Si utilizza tale vettore per trovare un terzo punto di supporto $V3$ sulla superficie

di $B - A$ come mostrato in figura 4.27 (a). Se l'origine cade al di fuori della retta di supporto parallela al segmento che unisce $V1$ e $V2$ e passante per $V3$, allora l'origine sarà esterna a $B - A$ e quindi l'algoritmo riporterà la non intersezione. Nel caso in esempio, l'origine cade all'interno della retta di supporto quindi l'algoritmo continua.

Si consideri il triangolo formato dai tre punti $V1$, $V2$ e $V3$ mostrato in figura 4.27 (b). Si noti che per costruzione la retta che unisce $V0$ all'origine passa attraverso la base $\overline{V1V2}$ del triangolo siccome quest'ultima è un portale. Di conseguenza la retta dovrà attraversare uno degli altri due lati del triangolo. Per determinare da quale lato esca si costruisce un segmento che unisce il nuovo punto di supporto $V3$ a $V0$. Se l'origine giace su uno dei due semipiani di questo segmento allora la retta uscirà attraversando il lato del triangolo che si trova sullo stesso semipiano.

Infine il lato del triangolo che viene selezionato diverrà il nuovo portale e il vecchio punto di supporto $V2$ verrà abbandonato. L'algoritmo ripeterà quindi il processo di perfezionamento del portale usando il nuovo portale mostrato in figura 4.27 (c). In questo esempio, alla seconda iterazione, l'origine giace all'interno del nuovo portale e di conseguenza l'algoritmo terminerà riportando una intersezione.

4.8.4 MPR in 3D

Come dimostrato, la versione 2D del MPR costruisce una semiretta che passa per un punto interno a $B - A$ e l'origine. Successivamente trova il portale che tale semiretta attraversa e lo rifinisce muovendolo sempre più vicino alla superficie di $B - A$ finché l'origine non si trovi all'interno del portale o esternamente a una linea di supporto sulla superficie della forma. Nel caso 3D il procedimento è analogo, la differenza è che in questo caso i portali costruiti non saranno più segmenti ma triangoli e si cercherà l'eventuale lato di uscita della retta dal tetraedro e non dal triangolo. Per la descrizione completa del *Minkowski Portal Refinement* in 3D si rimanda all'appendice B.

4.9 Sovrapposizione iniziale con sfere e cuboidi

Una ulteriore ottimizzazione consiste di verificare prima l'intersezione di sfere e cuboidi che contengano al loro interno gli oggetti che ci interessano. L'intersezione tra sfere viene verificata guardando la distanza dei due centri di massa dei due oggetti considerati. Nel caso questa fosse minore di due volte l'estensione massima dell'oggetto iniziale si avrà una intersezione. Nel caso dei cilindri l'estensione massima e quindi il raggio della sfera sarà:

$$r_{sfera} = \sqrt{r_{cil}^2 + (h_{cil}/2)^2} \quad (4.8)$$

Mentre l'interazione tra cuboidi viene calcolata utilizzando lo stesso algoritmo per l'intersezione delle molecole, quindi SAT o *XenoCollide*, nel caso di cilindri il cuboide avrà dimensione $\mathbf{L} = (2r_{cil}, 2r_{cil}, h_{cil})$. Nel caso non ci fosse intersezione tra sfere o cuboidi allora non servirà verificare l'intersezione tra i poliedri

completi, permettendo così di risparmiare tempo computazionale. Nel caso ci fosse intersezione tra sfere allora si dovrà verificare l'intersezione tra cuboidi e nel caso ci fosse intersezione anche di questi allora si verificherà l'intersezione tra poliedri.

Nel caso dei semidischi si costruisce la sfera centrandola a metà raggio e a metà altezza del semidisco e avrà raggio pari a:

$$r_{sfera} = \sqrt{h^2/4 + 5r^2/4} \quad (4.9)$$

Mentre il cuboide avrà una faccia in comune con il semidisco, ovvero quella del lato tagliato, e avrà dimensione $\mathbf{L} = (2r_{semicil}, r_{semicil}, h_{semicil})$. Questa ottimizzazione verrà utilizzata in cascata alle *linked cell list*, infatti l'intersezione tra sfere e cuboidi verrà effettuata solo per cilindri, o semidischi, in celle adiacenti.

4.10 Analisi dei tempi di esecuzione

Si effettua un confronto dei tempi di esecuzione dei due algoritmi di intersezione, SAT e *XenoCollide*. Del primo se ne confrontano tre versioni per verificare che le ottimizzazioni del paragrafo 4.7.5 abbiano apportato un miglioramento. Si confronta la versione senza alcuna ottimizzazione, indicata con SAT-3, la versione in cui si proietta solo su direzioni non già considerate e quindi si sfruttano le simmetrie dell'oggetto considerato, indicata con SAT-2. Infine la versione con tutte le ottimizzazioni possibili indicata con SAT, dove si utilizzano tutte le ottimizzazioni indicate nel paragrafo 4.7.5.

Il motivo per cui si è implementato il SAT, oltre che per confrontare i tempi di esecuzione dei due algoritmi, è stato per verificare la stabilità dell'algoritmo *XenoCollide* [22] reso disponibile dall'autore su github.com [16]. Come descritto negli scorsi paragrafi, l'algoritmo *Xenocollide* necessita della definizione di una tolleranza sulla distanza tra portale e superficie tonda. Avendo quindi dei criteri di terminazione numerici è necessario testare la stabilità di tale algoritmo. Per fare un confronto sulle risposte dei due algoritmi si è creato un programma dove in una box di dimensioni limitate con un cilindro nell'origine viene inserito un secondo cilindro con posizione e orientazione casuali. Si verifica quindi l'intersezione tra i due cilindri. Questo processo è simile a quello per il calcolo del volume escluso.

La verifica dell'intersezione tra i due cilindri viene fatta sia con l'algoritmo SAT che con l'algoritmo *XenoCollide*, verificando che i due risultati non siano differenti. Il test è stato fatto per un totale di 10^{12} inserimenti e non è stata riscontrata alcuna differenza nell'output dei due algoritmi di intersezione, confermando, in conclusione, che *XenoCollide* è un algoritmo stabile.

Verificata la stabilità dell'algoritmo *XenoCollide* rimane da confrontare i tempi di esecuzione tra gli algoritmi e scegliere il più rapido. Nel confronto si utilizza una simulazione in *ensemble* NVT con $N = 750$ cilindri e si varia il numero di vertici che approssimano i cilindri. I risultati per i due algoritmi nelle loro differenti versioni sono mostrati in figura 4.28.

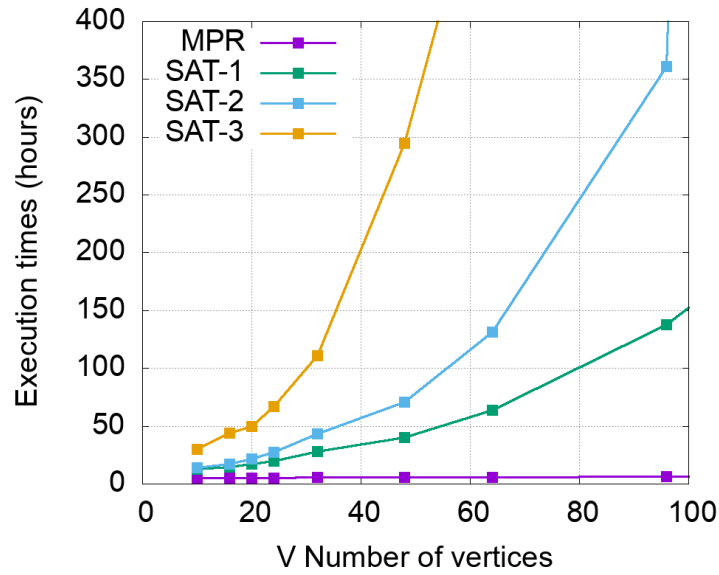


Figura 4.28. Confronto dei tempi di esecuzione per i due differenti algoritmi SAT e *XenoCollide*. SAT-1 indica la versione più veloce con tutte le ottimizzazioni. Tempi per completare una simulazione NVT di cilindri duri approssimati a poliedri con V .

Come mostrato in figura 4.28, *XenoCollide* risulta notevolmente più veloce. In particolare nella zona in cui si effettueranno le prossime simulazioni, ovvero $V = 18$ e $V = 32$ vertici, risulta dalle tre alle quattro volte più veloce. Anche nel caso in cui $V = 8$, ovvero l'intersezione tra cuboidi, risulta più veloce *XenoCollide* quindi non è neanche vantaggioso creare un algoritmo misto dove nel caso di intersezioni tra oggetti con pochi vertici utilizzi SAT e nell'altro caso *XenoCollide*.

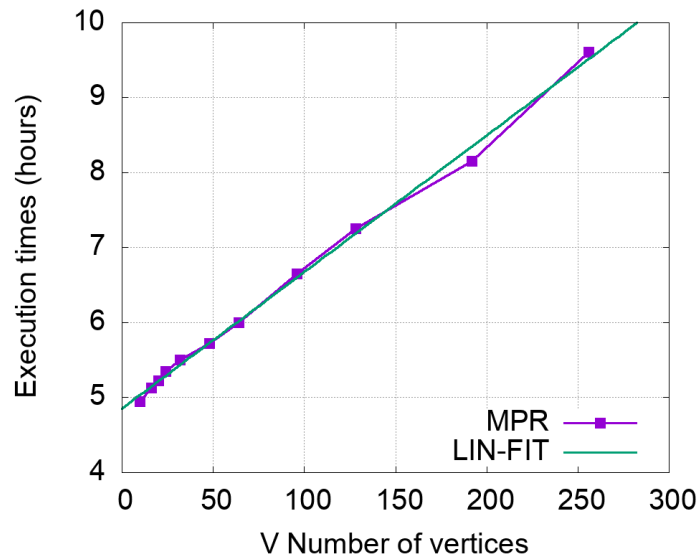


Figura 4.29. Tempi di esecuzione per l'algoritmo *XenoCollide* al variare del numero di vertici del poliedro.

Si noti che mentre l'algoritmo SAT presenta un andamento cubico rispetto al numero di vertici, *XenoCollide* ha un andamento lineare, come mostrato nell'immagine 4.29, quindi se si cerca di simulare oggetti con un numero di vertici alto la disparità tra i due algoritmi è notevole. In conclusione si decide di utilizzare per le simulazioni l'algoritmo *XenoCollide* visto il notevole vantaggio che porta in termini di tempi computazionali. Lo si utilizzerà in tutte le simulazioni nei prossimi capitoli a meno che diversamente specificato.

4.11 Funzione di distribuzione a coppie $g(\mathbf{r})$

È possibile distinguere le fasi in cui il sistema si trova andando a calcolare la funzione di distribuzione a coppie $g(\mathbf{r})$, che definisce la correlazione spaziale dello stato. La funzione definisce la probabilità di trovare una particella nella posizione \mathbf{x}_i considerata una particella nell'origine. La funzione è definita matematicamente come:

$$g(\mathbf{r}) = \frac{1}{\rho N} \left\langle \sum_{i=1}^N \sum_{j \neq i} \delta(\mathbf{r} - (\mathbf{r}_i - \mathbf{r}_j)) \right\rangle \quad (4.10)$$

Dove con ρ è la densità del sistema, N il numero di particelle, \mathbf{r}_i la posizione della i -esima particella e $\delta(\mathbf{r})$ è la funzione delta di Dirac [36]. La funzione di distribuzione a coppie deriva dal rapporto della densità a due particelle e il quadrato della densità del sistema:

$$g(\mathbf{r}) = \frac{\rho^{(2)}(\mathbf{r})}{\rho^2} \quad (4.11)$$

dove la densità a due particelle, ovvero la probabilità di trovare una particella in \mathbf{r}_1 data una particella in posizione \mathbf{r}_2 , è definita come:

$$\rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = \frac{N(N-1)}{Z} \int e^{-\beta U(\mathbf{r}_1, \dots, \mathbf{r}_N)} d\mathbf{r}_3 \dots d\mathbf{r}_N \quad (4.12)$$

Per calcolare la funzione di distribuzione a coppie nel caso delle simulazioni fatte nella tesi è necessario fare una precisazione, nelle simulazioni vengono utilizzate le condizioni periodiche al bordo e di conseguenza la distanza massima tra due particelle sarà $L/2$, dove L è la dimensione della box. Di conseguenza per procedere al calcolo della funzione di distribuzione a coppie si considera, per ogni particella, solo le particelle nel quadrato, con lato $L/2$. Si procederà quindi a costruire delle mappe di calore bidimensionali, ovvero per ogni particella si proiettano le posizioni delle particelle vicine su un piano, scelto in base alla direzione su cui si è interessati a verificare che ci siano periodicità, e si contano il numero di occorrenze su le possibili posizioni. La mappa di calore sarà un grafico delle occorrenze in funzione delle posizioni. Nel caso l'asse nematico sia parallelo all'asse z , saranno di interesse $g(0, y, z)$, $g(x, 0, z)$ e $g(x, y, 0)$. Queste saranno le correlazioni proiettate su diversi piani calcolate attraverso un programma in C++ che media su un numero di configurazioni di input. La mappa di calore verrà poi realizzata con *gnuplot*. In generale la funzione di distribuzione a coppie mostra se il sistema ha qualche grado di ordinamento, infatti, come mostrato nel capitolo 5.4.1, nel caso di sistemi isotropi la

mappa di calore non presenterà ordinamenti mentre, nel caso di cristalli, presenterà degli spot nelle posizioni periodiche del cristallo.

4.12 Catena di polimeri

Si tratta brevemente la teoria delle catene di polimeri per poter definire delle grandezze che descrivano i sistemi, con una attenzione particolare alla *persistence length* che sarà necessaria per adattare il modello alle caratteristiche sperimentali.

4.12.1 Catena a incastro libero

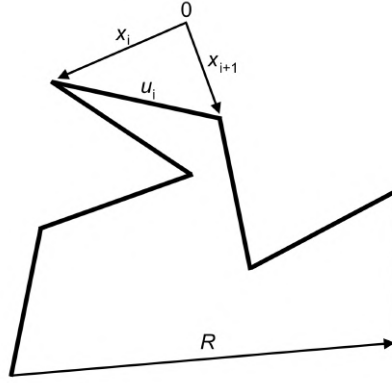


Figura 4.30. Rappresentazione schematica della catena a incastro libero.

Si descrive inizialmente un modello polimerico molto semplice, una catena ad incastro libero, composta da una sequenza di N segmenti rigidi, ciascuno di lunghezza l e in grado di puntare in qualsiasi direzione indipendentemente l'una dall'altra, come mostrato in figura 4.30. Per caratterizzare la conformazione della catena, consideriamo il vettore *end-to-end* R e calcoliamo il quadrato medio $\langle R^2 \rangle$ calcolando la media su tutte le possibili configurazioni [21]. Si indica il raggio vettore dell'inizio dell' i -esimo segmento con x_i e quello della sua fine con x_{i+1} . Inoltre, si introducono i vettori di legame, definiti come $u_i = x_{i+1} - x_i$. Pertanto, il vettore *end-to-end* R può essere scritto come:

$$R = \sum_{i=1}^N u_i \quad (4.13)$$

Mentre $\langle R^2 \rangle$:

$$\langle R^2 \rangle = \left\langle \left(\sum_{i=1}^N u_i \right)^2 \right\rangle = \sum_{i=1}^N \langle u_i^2 \rangle + 2 \sum_{i=1}^j \sum_{j=i+1}^N \langle u_i u_j \rangle \quad (4.14)$$

Siccome le direzioni dei segmenti in una catena liberamente articolata non sono correlate e l'angolo tra i vettori u_i e u_j ($i \neq j$) assume, con uguale probabilità, qualsiasi valore da 0 a 2π , $\langle u_i u_j \rangle = l^2 \langle \cos \theta_{ij} \rangle = 0$. Inoltre, $\langle u_i^2 \rangle = l^2$, così che:

$$\langle R^2 \rangle = Nl^2 \quad (4.15)$$

Quindi, per una catena lunga ($N \gg l$), la dimensione media di una macromolecola $\langle R^2 \rangle^{1/2} \approx N^{1/2}l$ è molto inferiore alla lunghezza totale Nl misurata lungo il contorno della catena polimerica. Questo implica che nell'insieme delle conformazioni che la catena liberamente articolata assume nel processo di moto termico, le conformazioni allungate (quasi dritte) costituiscono una frazione minore. La maggioranza assoluta delle conformazioni della catena è rappresentata da linee fortemente ripiegate nello spazio.

4.12.2 Flessibilità di una catena di polimeri

Nell'esempio precedente la catena è completamente flessibile, ovvero la persistence length in un unità di monomeri è 0. La flessibilità della catena a incastro libero è causata da connessioni libere di ruotare tra segmenti rigidi. Questo meccanismo di flessibilità ad incastro libero è facile da descrivere ma molto difficile da realizzare nella pratica; si osserva in pochissime sostanze reali. Tuttavia, tutte le catene polimeriche sufficientemente lunghe sono abbastanza flessibili. In particolare, si consideri il caso limite più sfavorevole, ovvero che la conformazione a catena lineare corrisponda al minimo assoluto di energia e che l'energia di eccitazione termica produca solo piccole deformazioni della struttura. Per piccole deformazioni, la struttura atomica di una molecola può essere considerata come una classica costruzione elastica, che nel caso del polimero è approssimata da un filamento sottile, elastico, omogeneo che obbedisce alla legge di Hooke sotto deformazione angolare. Un tale modello di catena polimerica è chiamato persistente.

L'esame del modello persistente mostra che anche piccoli piegamenti di fluttuazione delle sue sezioni portano all'avvolgimento totale di una catena sufficientemente lunga, perché diverse sezioni si piegano su lati diversi. Sebbene la flessibilità vera e propria sia un attributo generico di tutte le macromolecole, il meccanismo di flessibilità può essere diverso per i diversi polimeri. Molti polimeri a catena rigida e macromolecole elicoidali sono caratterizzati dal meccanismo di flessibilità persistente, cioè da una flessibilità uniforme su tutta la lunghezza. Per la macromolecola di DNA a doppia elica, ad esempio, il meccanismo di flessibilità persistente fornisce un'approssimazione soddisfacente [21].

Il meccanismo di flessibilità persistente è in un certo senso il più fondamentale, perché interviene ogni volta che le oscillazioni termiche sono armoniche. L'esistenza di una flessibilità persistente può essere dimostrata anche senza alcuna conoscenza della natura chimica dei polimeri. Al contrario, tutti gli altri meccanismi di flessibilità sono causati da vari effetti anarmonici per cui la realizzazione di un meccanismo specifico dipende dalla struttura chimica individuale del polimero.

4.12.3 *Persistence Length*

La correlazione direzionale di due segmenti di una macromolecola diminuisce esponenzialmente con la lunghezza della catena che li separa. Se si riconsidera l'equazione 4.14, per un modello di catena diverso da quello a giunzione libera, $\langle u_i u_j \rangle \neq 0$, perché le direzioni dei diversi segmenti di catena sono correlate. Se $\langle u_i u_j \rangle \approx \langle \cos \theta_{ij} \rangle$,

questa correlazione, che determina un grado di flessibilità della catena, può essere espressa qualitativamente dal coseno medio dell'angolo tra i diversi segmenti del polimero. Si introduce quindi la grandezza $\langle \cos \theta(s) \rangle$, la media del coseno dell'angolo tra i segmenti di catena separati da s .

Questa funzione di s per molti modelli di catena polimerica possiede la proprietà della moltiplicatività, se la catena ha due sezioni vicine con lunghezze s e s' , allora:

$$\langle \cos \theta(s + s') \rangle = \langle \cos \theta(s) \rangle \langle \cos \theta(s') \rangle \quad (4.16)$$

La funzione che presenta questa proprietà è la funzione esponenziale, quindi:

$$\langle \cos \theta(s) \rangle = e^{-s/l_p} \quad (4.17)$$

Il fattore preesponenziale è pari a uno, siccome $\cos(\theta(s = 0)) = 1$ e l_p sono delle costanti per ogni dato polimero. La costante l_p è la caratteristica fondamentale della flessibilità del polimero ed è chiamata la *persistence length* del polimero. La *persistence length* è una proprietà meccanica che quantifica la rigidità alla flessione di un polimero. In generale, i polimeri che sono più corti della *persistence length*, si comportano come rigidi, mentre per polimeri più lunghi della *persistence length* l_p i polimeri possono essere descritti statisticamente come un random walk tridimensionale. Formalmente, la *persistence length* è definita come la lunghezza oltre la quale si perdono le correlazioni tra le tangenti ai polimeri ovvero le fluttuazioni di flessione cancellano la memoria della direzione della catena.

Per chiarire il senso fisico dell'equazione 4.17 si consideri l'angolo θ tra due tratti vicini della catena, separati da un tratto di catena la cui lunghezza è molto minore della *persistence length*: $s \ll l_p$. In queste condizioni, $\cos(\theta(s = 0)) \approx 1$. Ciò implica che la sezione della catena più corta di l_p non mostra alcuna flessibilità e si comporta essenzialmente come un'asta rigida ($\theta \approx 0$). Nel caso limite opposto $s \gg l_p$, quando le sezioni considerate sono separate da un'ampia porzione della catena, l'equazione 4.17 avrà $\cos(\theta(s = 0)) \approx 0$. Quindi, l'angolo θ assume qualsiasi valore compreso tra 0 e 2π , e la flessibilità della catena porta alla pressoché totale indipendenza delle sezioni macromolecolari separate da una distanza pari o superiore alla lunghezza persistente. In altre parole, la memoria della direzione della catena si perde su una distanza paragonabile alla lunghezza persistente.

In generale il calcolo della *persistence length* viene effettuato andando a calcolare un *fit* esponenziale della funzione correlazione spaziale C_0 , definita come:

$$C_0(s) = \langle u_i \cdot u_{i+s} \rangle \quad (4.18)$$

Dove, se $i = 0$, u_i indicherà il primo monomero del filamento; Mentre $\langle \dots \rangle$ indica la media su tutte le configurazioni e su tutti i possibili valori di i . Questa quantità sarà pari all'equazione 4.17, quindi andando a effettuare un fit esponenziale è possibile ricavare l_p . Se il polimero considerato presenta una *persistence length*

alta, e quindi l'angolo θ tra due monomeri contigui è vicino a zero, è possibile fare un'approssimazione in Eq. 4.17 determinando una semplice relazione tra la *persistence length* e il momento secondo della distribuzione associata all'angolo θ . Considerando una sezione corta del filamento, $s \ll l_p$, si avrà infatti:

$$\langle \cos \theta(s) \rangle \approx 1 - \frac{s}{l_p} \quad (4.19)$$

E, siccome l'angolo $\theta(s)$ è piccolo, $\cos(\theta(s)) \approx 1 - \theta^2(s)/2$, in conclusione:

$$\langle \theta^2(s) \rangle \approx \frac{2s}{l_p} \longrightarrow l_p \approx \frac{2s}{\langle \theta^2(s) \rangle} \quad (4.20)$$

Quindi andando a misurare l'angolo tra monomeri primi vicini è possibile calcolare la *persistence length* come $l_p \approx 2/\langle \theta^2(s) \rangle$. Questo è un vantaggio notevole rispetto la stima tramite fit esponenziale siccome in quel caso è necessario simulare filamenti lunghi circa la *persistence length*. Il costo computazionale di far evolvere un filamento molto lungo è alto, perché è necessario attendere che perda memoria della configurazione iniziale o di quella della misura precedente. Nel caso della stima attraverso l'angolo tra primi vicini θ è possibile simulare filamenti nettamente più corti, in particolare sono sufficienti filamenti lunghi 5 monomeri che saranno nettamente più veloci da far evolvere.

4.12.4 Stima della *persistence length* per sfere dure con patch attrattive

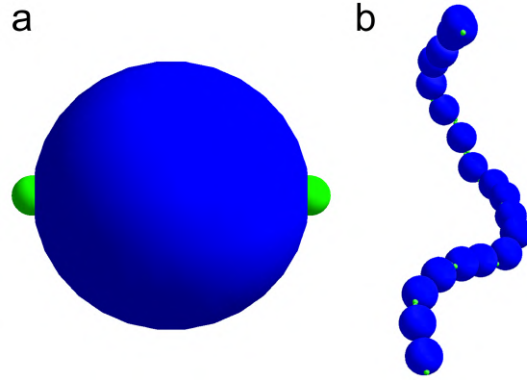


Figura 4.31. (a) Vista laterale della sfera con due patch. (b) Filamento composto da venti sfere di raggio pari a uno e $r_{patch} = 0.15$

Per verificare l'approssimazione mostrata nell'equazione 4.20 si procede a creare un sistema semplificato composto da sfere dure con due patch attrattive inserite ai due poli opposti della sfera, come mostrato in figura 4.31. Così facendo è possibile far evolvere il sistema più velocemente, anche con filamenti lunghi, e di stimare la *persistence length* tramite il fit esponenziale. Si calcola la $C_0(s)$ al variare del raggio della patch attrattive, nell'immagine 4.32 si mostrano i valori ottenuti per filamenti lunghi 20 monomeri. L'andamento rimane esponenziale e la pendenza aumenta all'aumentare del raggio delle patch.

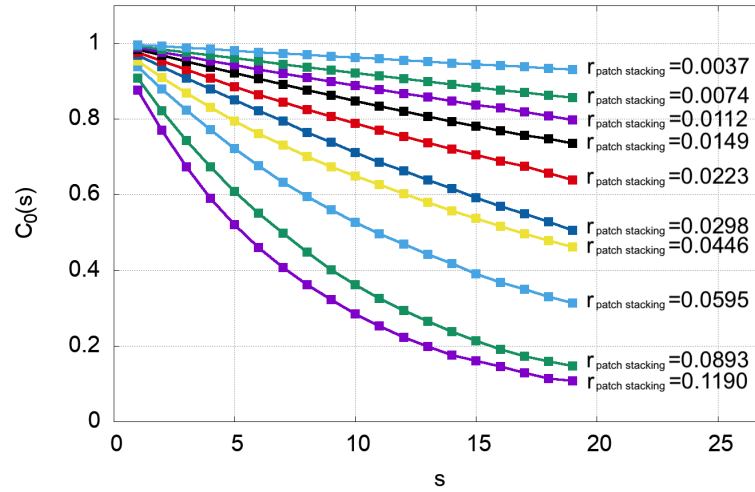


Figura 4.32. Calcolo della funzione correlazione spaziale $C_0(s)$ dove s è la distanza tra i due monomeri considerati.

Si procede quindi a stimare la *persistence length* anche utilizzando l'equazione 4.20, in questo caso si media il θ^2 tra primi vicini. La misura viene effettuata facendo passare un numero sufficiente di passi Monte Carlo per perdere la correlazione con la misurazione precedente. Si confrontano le due stime ottenute per verificare che siano compatibili, il confronto è mostrato in figura 4.33.

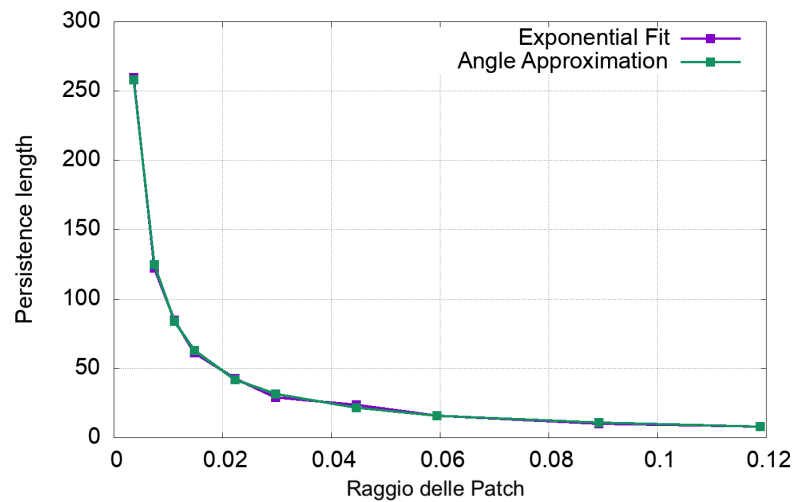


Figura 4.33. Grafico della *persistence length* in funzione del raggio delle patch. Le sfere hanno raggio pari a $r = 1$.

Come si nota c'è un ottimo accordo tra le due stime, inoltre bisogna notare che la stima tramite fit esponenziale converge al valore più lentamente. Sono necessari anche filamenti lunghi per un fit preciso ma di conseguenza i passi Monte Carlo tra due misurazioni devono aumentare, aumentando, notevolmente, i tempi computazionali.

Inoltre si mostra che le stime della persistence length ottenute sono in accordo con quelle riportate in [51], l'articolo riporta una *persistence length* compresa tra [10 : 20] e per le stesse dimensioni delle patch con entrambi i metodi si è ottenuto circa 11. Di conseguenza si decide di utilizzare in seguito la stima tramite l'angolo tra primi vicini anche a causa del fatto che la stima tramite fit per filamenti costituiti da semidischi è molto costosa in termini di tempi computazionali.

4.12.5 *Persistence length* di filamenti di semidischi

Come discusso nel paragrafo 4.1.4, la dimensione delle patch di *stacking* e di *pairing* e la loro posizione permettono di modificare i parametri strutturali del filamento creato per adattarlo ai parametri del filamento di DNA sperimentale. In particolare è possibile andare a modificare la *persistence length* e la lunghezza totale del filamento. Sperimentalmente si misura una *persistence length* di circa 50 nm [18], mentre, come detto nel paragrafo 2.3, la lunghezza di un filamento di DNA composto da dodici basi accoppiate è di circa 4 nm.

Per stimare la *persistence length* nel caso di filamenti di semidischi, si utilizza il metodo con i primi vicini e quindi l'equazione 4.20. Si calcola il versore medio tra due vettori normali a due semidischi accoppiati secondo legame di *pairing* e poi si valuta l'angolo θ tra i vettori medi calcolati per le varie basi accoppiate del filamento. Per convergere al valore medio più velocemente si simulano filamenti lunghi 5 basi, siccome questi perdono memoria della configurazione precedente più velocemente. Tuttavia qui c'è una precisazione da fare, l'angolo θ non va valutato tra tutte le basi siccome quello tra un estremo del filamento e la base successiva è viziato dal fatto che l'estremo può assumere configurazioni che rendono impossibile il proseguimento del filamento. Di conseguenza nel filamento di 5 basi vanno considerati solamente i due legami centrali, e, in ogni caso, se si utilizzano filamenti più lunghi nella simulazione, i legami degli estremi non andranno considerati. Questo metodo tiene conto quindi solo degli angoli tra dischi che potrebbero far parte di un filamento teoricamente infinito.

In prima analisi si stima la *persistence length* fissando la grandezza delle patch di *stacking* e di *pairing* e variando la distanza delle patch di *stacking* dal centro del semidisco. Si varia questa quantità da $0.15 \cdot r_{\text{semidisco}}$, perché se inferiore le patch potrebbero legarsi con il semidisco già legato dal *pairing*, a $0.5 \cdot r_{\text{semidisco}}$. Le patch di *pairing* vengono scelte di un raggio pari a $r_{\text{pairing}} = 0.05$ perché così si favorisce la formazione di dischetti composti da due semidischi e si evitano legami di *pairing* che formino dischi piegati che impediscano la continuazione del filamento. Nell'immagine 4.34 viene riportato il grafico ottenuto, andando a avvicinare le patch di *stacking* al centro del cilindro la *persistence length* diminuisce come ci si aspettava.

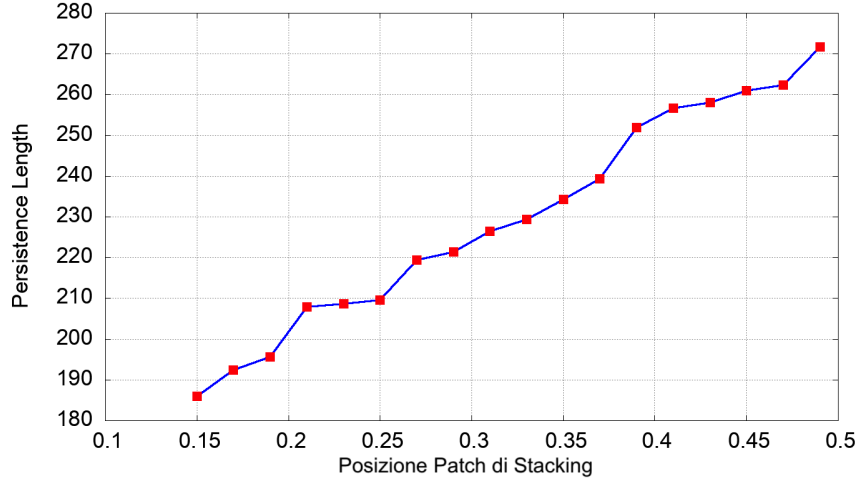


Figura 4.34. Grafico della *persistence length* in funzione della posizione della patch di *stacking*. Sulle ascisse si indica la distanza della patch dal centro del semidisco. Patch di *pairing* di raggio $r_{pairing} = 0.05$.

Una precisazione da fare è che le patch di *stacking* hanno un raggio massimo possibile, questo è determinato dal fatto che le patch di *stacking* dello stesso semidisco non debbano toccarsi tra loro. Questa richiesta è necessaria perché, nel caso in cui la limitazione di un legame per patch sia attiva, non sarebbe concesso ai semidischi di legarsi con le patch di *stacking* molto vicini tra loro. Se la limitazione fosse disattivata allora si avrebbe un conteggio doppio nell'energia. Nel caso di filamenti con rotazione simile al DNA le due patch di *stacking* dello stesso filamento sono ruotate di $\theta = \pi/5$, di conseguenza questo permette di aumentare il raggio delle patch, in particolare il limite, in funzione dell'angolo θ è dato dalla formula:

$$2r_{semidisco} = \sqrt{h_{semidisco}^2 + (2k \sin(\theta/2))^2} \quad (4.21)$$

dove con k si indica la distanza della patch rispetto il centro del semidisco. Di conseguenza andando ad allontanare dal centro nel posizionamento della patch, quindi aumentando k , è possibile aumentare la dimensione delle patch di *stacking*.

Si decide di posizionarsi a $k = 0.35 \cdot r_{sd}$, evitando così il fenomeno dello sdoppiamento dei filamenti presentato nello scorso capitolo nella figura 4.5, e per questo posizionamento si analizza come varia la *persistence length* andando a variare la dimensione delle patch *stacking*. I risultati sono mostrati nel grafico in figura 4.35.

Scegliendo quindi un raggio delle patch pari a $r_{stacking} = 0.14$ si ottiene una *persistence length* di circa 150 basi, questa sarà la scelta che si utilizzerà per simulare filamenti di DNA. Per calcolare la lunghezza totale di un filamento composto da 24 basi si procede prima calcolando il punto medio tra due semidischi legati con patch di *pairing*, ponendo attenzione alle condizioni periodiche al bordo, successivamente si calcolano le distanze tra i punti medi di due basi legate di *stacking*. Sommando tutte queste distanze, più una volta l'altezza di un semidisco per tener conto delle due mezze altezze non calcolate degli estremi del filamento, si ottiene la lunghezza del filamento.

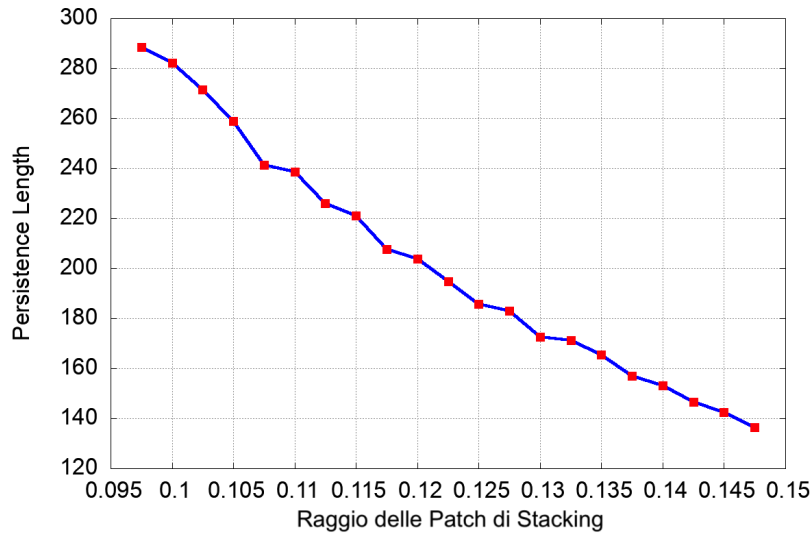


Figura 4.35. Grafico della *persistence length* in funzione del raggio della patch di *stacking*.

Sperimentalmente si misura una lunghezza del filamento di DNA, lungo 12 basi, di circa 4 nm. Si effettuano delle simulazioni con filamenti di 24 basi totali con legami permanenti per misurare le lunghezze. Dopo aver effettuato 50000 passi d'equilibratura, viene effettuata una misura con 30000 passi MC per ogni filamento. Si procede studiando la lunghezza del filamento al variare dell'altezza del semidisco, il risultato è mostrato in figura 4.36.

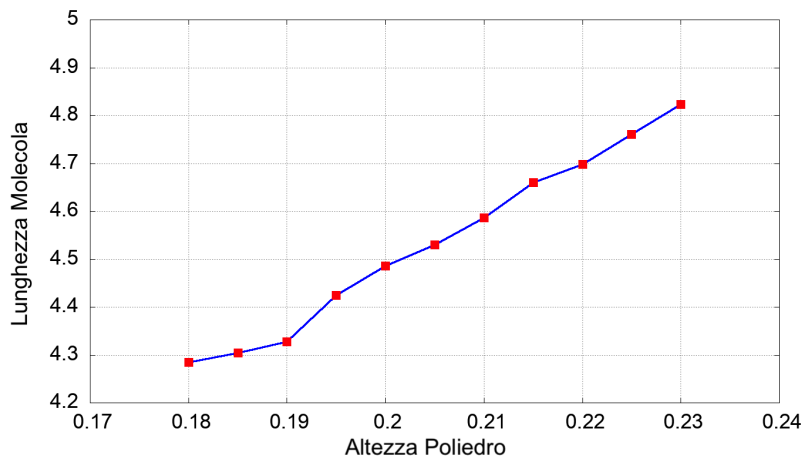


Figura 4.36. Grafico della lunghezza del filamento in funzione dell'altezza dei semidischi che lo compongono. Il raggio della patch di *stacking* è fissato a $r_{stacking} = 0.14$.

Si effettua lo stesso studio al variare del raggio della patch di *stacking*, i risultati sono mostrati in figura 4.37.

Riconsiderando il vincolo dato dall'equazione 4.21 e che per avere una *persistence length* di 150 basi è necessaria una patch di *stacking* di raggio $r_{stacking} = 0.14$ si

decide di utilizzare dei semidischi con altezza pari a $h = 0.2$, dando precedenza ad avere una flessibilità del filamento corretta. Così facendo si otterrà una lunghezza del filamento di circa 4.4 nm , valore ragionevolmente prossimo a quello reale di 4 nm .

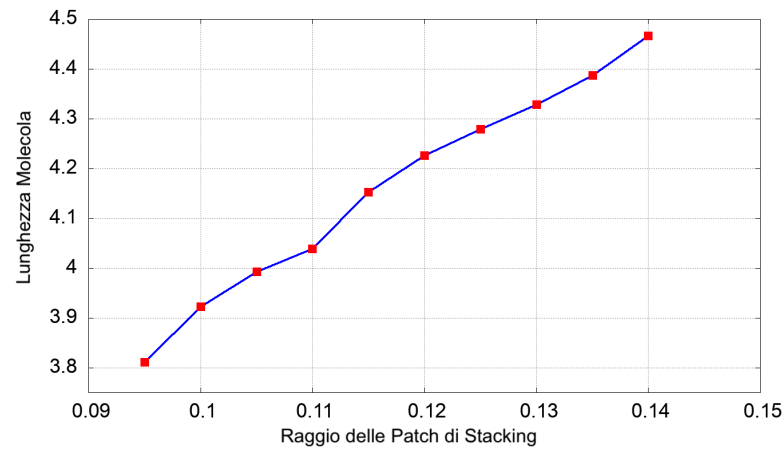


Figura 4.37. Grafico della lunghezza del filamento in funzione del raggio della patch di *stacking*. L'altezza del filamento è fissata a $h = 0.2$.

Capitolo 5

Simulazioni Monte Carlo di Cilindri

In questo capitolo si discutono le simulazioni effettuate per i cilindri, in particolare si verifica che l'approssimazione dei cilindri con un numero di vertici finiti, richiesto dagli algoritmi di intersezione, non modifichi il comportamento fisico del sistema per un certo numero minimo di vertici. Inoltre si studia l'adattamento da effettuare alle energie per passare da una simulazione con cilindri con una patch per base a una con due patch per base. Il passaggio a due patch per base permette di mantenere i legami nel momento in cui si sostituiscono i cilindri con filamenti di semidischi e ad avere una stima della temperatura ridotta da usare per i semidischi. Infine si studia la presenza delle fase nematica e colonnare per cilindri con 2 patch per base e *aspect-ratio* $X_0 = 2$.

5.1 Approssimazione a poliedri

In questo paragrafo si verifica che l'approssimazione a poliedri degli oggetti non modifichi la fisica della simulazione e si determina quale possa essere un buon numero di vertici per approssimare i poliedri così da ottenere un diagramma di fase in accordo con quello di cilindri esatti. A proposito, va considerato che il numero di vertici deve rimanere contenuto per evitare tempi troppo lunghi di simulazione, come visto nel paragrafo 4.10 utilizzando l'algoritmo *Xenocollide* il tempo di esecuzione è lineare con il numero di vertici. Si procede quindi a controllare l'equazione di stato per dei cilindri duri senza patch in un *ensemble* NPT, a una temperatura fissata per due diverse elongazioni dei cilindri. Per ottimizzare i tempi sono state fornite dal professore De Michele le configurazioni finali dei sistemi equilibrati per diversi punti dell'equazione di stato, si è quindi proceduto a sostituire i cilindri con la loro approssimazione poliedrica con un numero di vertici $V = 16, 24, 32$ e a far riequilibrare il sistema per altri due milioni di passi Monte Carlo.

In figura 5.1 e 5.2 si mostrano le equazioni di stato ottenute, come si può notare i cambi di fase del sistema sono mantenuti. Si trova solo una traslazione dell'equazione di stato verso frazioni di volume minori, questo spostamento è maggiore quando si approssimano i cilindri con un numero basso di vertici. In ogni caso si può notare che

per $V = 32$ vertici si ottiene una buona approssimazione, di conseguenza questo sarà il numero di vertici utilizzato nella tesi generalmente, se non diversamente indicato.

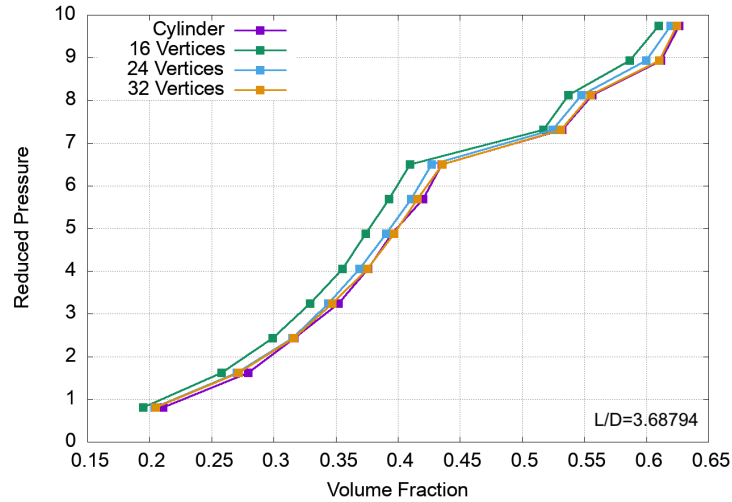


Figura 5.1. Equazione di stato dei cilindri con *Aspect-ratio* di 3.68794. Si confronta l'equazione di stato dei cilindri fornita dal professore De Michele con la loro approssimazione a poliedri con 16,24 e 32 vertici.

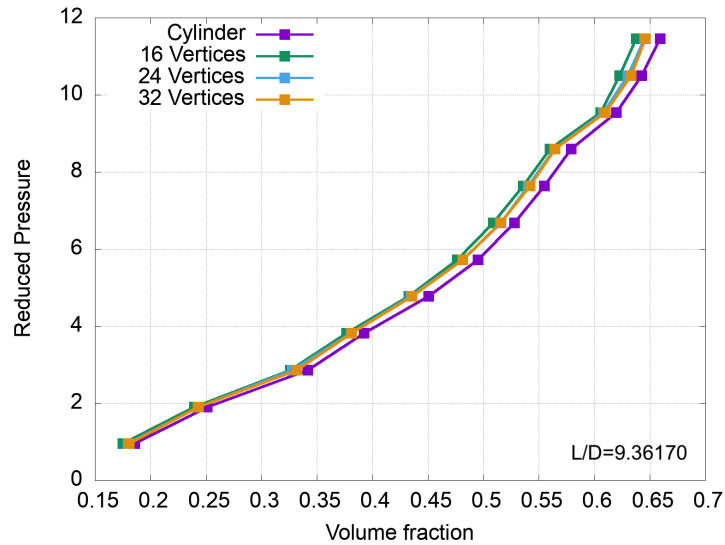


Figura 5.2. Equazione di stato dei cilindri con *Aspect-ratio* di 9.36170. Si confronta l'equazione di stato dei cilindri fornita dal professore De Michele con la loro approssimazione a poliedri con 24 e 32 vertici.

5.2 Stima della *Persistence Length* dei cilindri

Si cercano le fasi nematiche di cilindri con quattro patch attrattive, queste poi serviranno per andare a sostituire i cilindri con i semidichi. Per avere continuità

nella sostituzione si modifica posizione e dimensione delle patch dei cilindri così da avere la stessa flessibilità dei filamenti formati da cilindri e formati da semidischi. Si cerca di ottenere, in primo luogo, una fase nematica perciò si utilizza come punto di partenza del posizionamento delle patch quanto riportato in Ref. [36], tuttavia, avendo 2 patch per base, queste si posizionano alla stessa distanza dal centro come nel caso dei semidischi ovvero $k = 0.35 * r_{cil}$.

Un'altra differenza è la rotazione tra le patch delle due basi di $2\pi/5$, questo perché, come visto nel paragrafo 4.1.4, i semidischi compiono una rotazione completa nel filamento ogni dieci basi e visto che il filamento ha una lunghezza di dodici basi si avrà una rotazione completa delle patch più un angolo di $2\pi/5$.

Si stima la *persistence length* dei cilindri a 4 patch, l'obiettivo è quello di ottenere un valore pari a $p_l = 12.5$, visto che il cilindro andrà a simulare un filamento di 12 basi e quindi una PL di $p_l = 12.5 \cdot 12 = 150$ basi nucleotidiche. Si effettuano varie simulazioni per stimare la *Persistence Length* al variare della grandezza delle patch attrattive, i risultati sono mostrati in figura 5.3 e per ottenere una $p_l = 12.5$ si scelgono delle patch attrattive di raggio pari a $r = 0.285$.

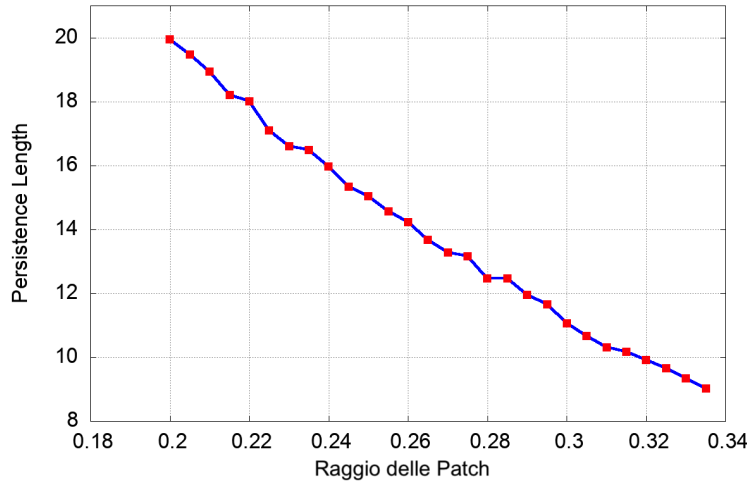


Figura 5.3. Stima della *Persistence Length* per cilindri con 4 patch attrattive e aspect-ratio $X_0 = 2$, grafico al variare della dimensione della patch.

5.3 Energia libera di *Stacking*

In questo paragrafo si discute come calcolare l'energia libera di *Stacking* [11], questa quantità sarà necessaria per il passaggio da cilindri a due patch a cilindri con quattro patch. Infatti nel momento in cui si sostituiscono i cilindri, nelle fasi nematiche o colonnari, con i dischi tagliati si passa ad avere una patch per base a due. Siccome non si vuole cambiare significativamente l'energia di legame è necessario imporre la stessa energia libera di *Stacking* nei due sistemi.

L'energia libera di *Stacking* è definita come:

$$\beta \Delta F_b = \ln \left[2 \frac{\Delta(T)}{v_d} \right] \quad (5.1)$$

dove:

$$\Delta(T) = \frac{1}{4} \left\langle \int_{V_b} [e^{-\beta V(\mathbf{r}_{12}, \Omega_1, \Omega_2)} - 1] d\mathbf{r}_{12} \right\rangle \quad (5.2)$$

Con \mathbf{r} si indica il vettore che unisce i due centri di massa della particella 1 e la particella 2, Ω_i è l'orientazione della i -esima particella e con $\langle \dots \rangle$ si rappresenta la media fatta su tutte le direzioni. v_d è il volume dell'oggetto considerato mentre V_b è il volume di bonding. Il calcolo di $\Delta(T)$, nel caso di patch attrattive come nella tesi, si semplifica perché basta calcolare la probabilità di ottenere un legame singolo e quella di ottenere un legame doppio per poi pesarle rispettivamente con $e^{\beta U_0}$ e $e^{2\beta U_0}$. Di conseguenza l'equazione per $\Delta(T)$ in 5.2 diventa:

$$\Delta(T) = \frac{1}{4N} (N_1(e^{\beta u_0} - 1) + N_2(e^{2\beta u_0} - 1)) \quad (5.3)$$

Dove con N si indica il numero di inserimenti totali, con N_1 e N_2 si indica il numero di volte in cui si hanno rispettivamente uno e due legami. Si procede quindi a realizzare un programma che inserisca un cilindro nell'origine e poi inserisca un secondo cilindro con posizione e orientazione casuale all'interno di una scatola con dimensioni tali da contenerli entrambi. Si verifica il numero di patch legate ad ogni inserimento, di cui se ne effettuano un totale di 5 miliardi. Si procede quindi a effettuare questa simulazione sia per i cilindri con una patch per base e due patch per base, nel secondo caso la giustificazione della posizione delle patch è stata discussa nel paragrafo precedente. Le patch avranno raggio pari a $r_{patch} = 0.285$ e nel caso di 4 patch, saranno a distanza dal centro del semidisco pari a $d = 0.35 \cdot r_{patch}$. I risultati degli inserimenti sono mostrati nella tabella 5.4.

Numero di patch	Cilindro			Lato della box	Numero di sovrapposizioni		Inserimenti totali
	Altezza	Diametro	Volume		Di una patch	Di due patch	
2	4.2	2.1	14.1761	10	333756	\	5.00E+09
4	4.2	2.1	14.1761	10	1295245	133902	5.00E+09

Figura 5.4. Tabella con i risultati del programma per il calcolo dell'energia libera di *stacking*. Il volume dei cilindri viene calcolato come volume dei poliedri siccome vengono approssimati con 32 vertici.

Effettuate le simulazioni si cerca di trovare la temperatura ridotta, definita come $T_{2patch}^* = k_B T_{2patch} / U_0$, tale per cui $\Delta F_b^{1patch}(T_{1patch}^*) = \Delta F_b^{2patch}(T_{2patch}^*)$, che verrà poi utilizzata nelle simulazioni dei dischi tagliati per avere la stessa energia libera di *Stacking*. Si calcola quindi l'energia libera di stacking ΔF_b utilizzando l'equazioni 5.1 e 5.3, l'andamento è mostrato in figura 5.5.

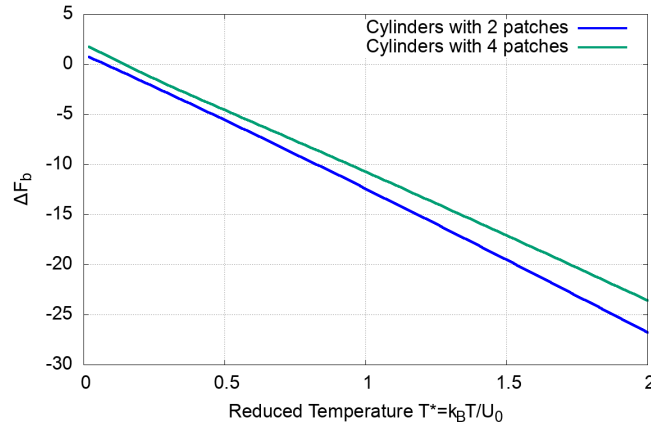


Figura 5.5. Andamento della ΔF_b in funzione della temperatura ridotta $T^* = k_B T / U_0$ per cilindri con 2 e 4 patch.

5.4 Fasi LC per cilindri con 4 patch

L'interesse della tesi è verificare la stabilità delle fasi liquido cristalline dei cilindri con patch attrattive una volta sostituiti con filamenti di semidischi. Di conseguenza è necessario creare delle configurazioni in fasi liquido cristalline equilibrate utilizzando cilindri con 4 patch. Le patch dei cilindri sono ruotate di $2\pi/5$ tra le due basi di modo che, dopo la sostituzione a semidischi, le patch attrattive sugli estremi rimangano nelle stesse posizioni e mantengano eventuali legami.

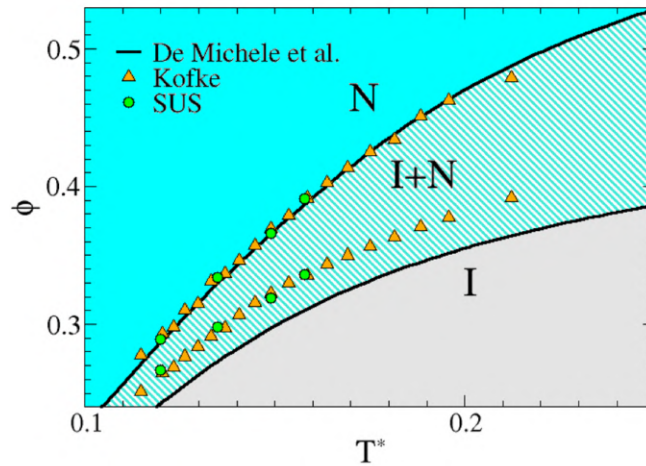


Figura 5.6. Diagramma della fase isotropica-nematica della frazione di volume ϕ in funzione della temperatura ridotta T^* . Le linee rappresentano la previsione teorica mentre i triangoli rappresentano le simulazioni MC. Immagine tratta da [36].

Come punto di partenza si utilizza i risultati riportati in Ref. [36], dove nell'immagine 5.6 viene rappresentata l'equazione di stato. Nel paper vengono simulati cilindri con una patch centrata su ogni base con *aspect-ratio* pari a $X_0 = L/D = 2$.

Si decide quindi di posizionarsi a $T^* = 0.12$ dove, nel caso la frazione di volume sia maggiore di 0.29, si ottengono fasi nematiche. Per effettuare il passaggio da due a quattro patch è necessario modificare la temperatura ridotta come descritto nello scorso paragrafo. Andando a interpolare la funzione, come mostrato in figura 5.7, si ottiene una temperatura ridotta $T_{4patch}^* = 0.18475$.

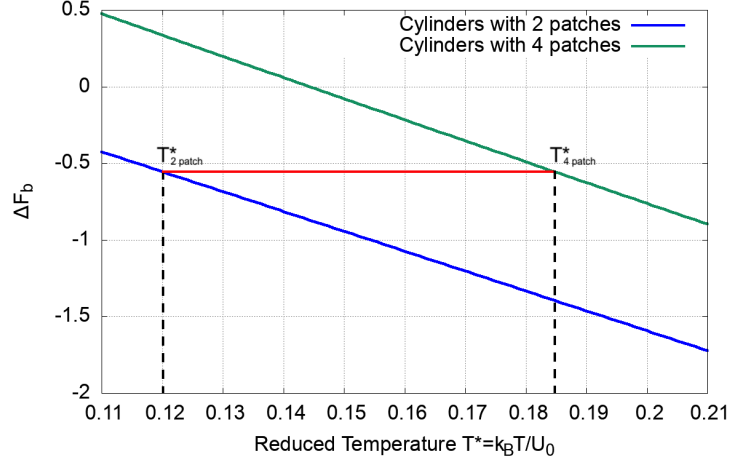


Figura 5.7. Calcolo della energia libera di stacking ΔF_b per cilindri con due e quattro patch, imponendola uguale per $T_{2patch}^* = 0.12$ è possibile trovare la temperatura ridotta nel caso con 4 patch $T_{4patch}^* = 0.18475$.

Un'altra precisazione sulle simulazioni è che andranno effettuate con un numero ridotto di molecole siccome, una volta completate, andranno sostituiti i cilindri con i filamenti di semidischi. Ogni cilindro sarà sostituito da 24 semidischi, perché andranno a simulare un filamento di DNA lungo circa 4 nm e quindi 12 coppie di nucleotidi. Di conseguenza il numero di molecole iniziale andrà moltiplicato per 24 e per limitare il tempo computazionale per le simulazioni con i semidischi si cerca di mantenere al minimo il numero di cilindri iniziali.

5.4.1 Fase nematica

Si procede, dunque, a effettuare una simulazione NPT a differenti pressioni e a T_{4patch}^* per cercare le fasi nematiche. Il numero di particelle viene contenuto a $N = 180$ per quanto detto in precedenza. Si evolve il sistema per 4 milioni di passi Monte Carlo, mentre la configurazione iniziale è cristallina. Per favorire la fase liquido cristallina nematica in direzione parallela all'asse \hat{z} si inserisce nei primi 10% dei passi un potenziale esterno [38] descritto dall'equazione:

$$U_{ext} = U \sum_{i=1}^N [1 - (\mathbf{u}_i \cdot \mathbf{z})^2] \quad (5.4)$$

Dove con \mathbf{u}_i indica l'orientazione della i -esima molecola, mentre U è il parametro che indica la forza del campo esterno.

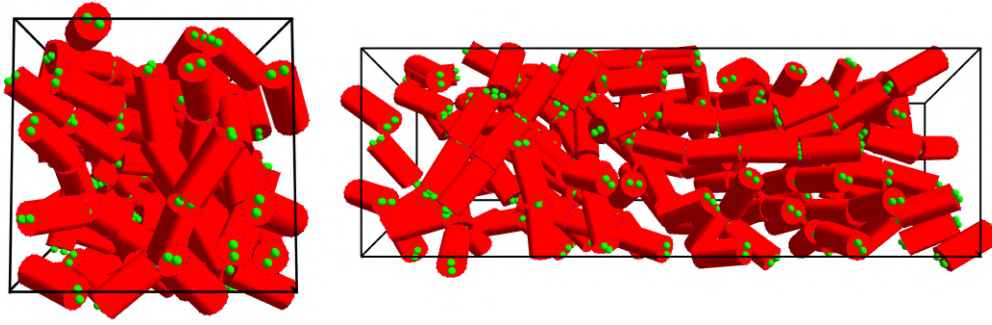


Figura 5.8. Vista frontale e in prospettiva della fase isotropa con 180 molecole a una pressione ridotta di $P^* = 0.45$ e una temperatura ridotta di $T^* = 0.18475$.

Si effettuano varie simulazioni con cilindri con *aspect-ratio* $X_0 = 2$ a quattro patch andando a variare la pressione. Compatibilmente all'articolo [36] si trova una transizione di fase intorno a una frazione di volume $\phi \approx 0.29$, al di sotto di tale concentrazione si trovano fasi isotrope. Le fasi isotrope, di cui una è rappresentata in figura 5.8, hanno una funzione di distribuzione radiale che non presenta periodicità in nessuna direzione, nell'immagine 5.9 (a) viene presentata la $g(x, y, 0)$, mentre nell'immagine 5.9 (b) viene mostrata la $g(0, y, z)$.

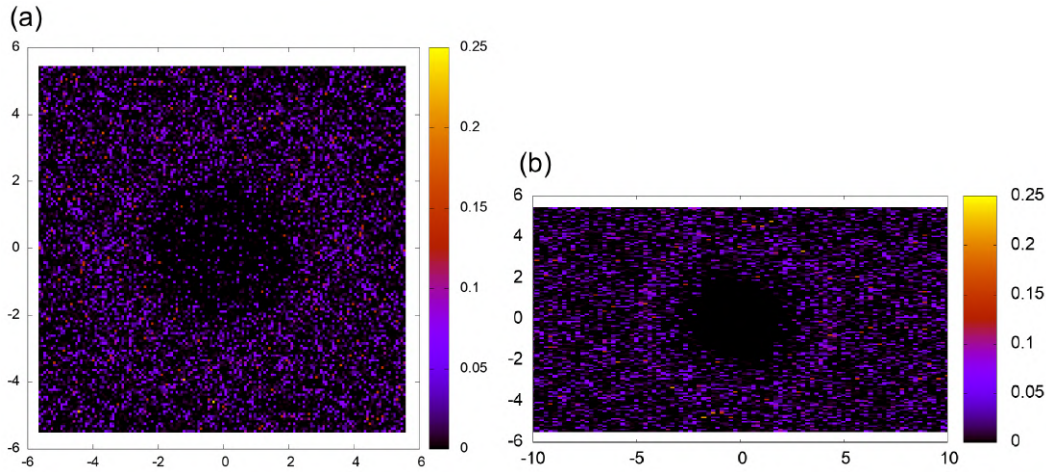


Figura 5.9. (a) Funzione di distribuzione radiale $g(x, y, 0)$ della fase isotropa. (b) Funzione di distribuzione radiale $g(0, y, z)$ della fase isotropa. Il sistema non presenta periodicità in nessuna direzione.

A partire da pressioni superiori a $P^* > 0.5$ si trovano fasi nematiche, infatti presentano un parametro d'ordine nematico $S > 0.6$. Andando a aumentare ulteriormente la pressione la fase nematica diventa più compatta e anche il parametro d'ordine nematico aumenta. Una delle prime fasi nematiche trovate è mostrata in figura 5.10, ha un parametro d'ordine nematico pari a $S = 0.764$, con l'asse nematico diretto parallelamente all'asse \hat{z} .

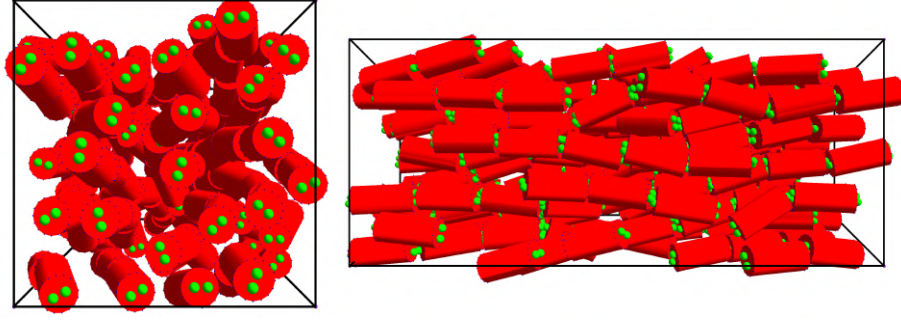


Figura 5.10. Vista frontale e laterale della fase nematica con 180 molecole a una pressione ridotta di $P^* = 0.55$ e una temperatura ridotta di $T^* = 0.18475$.

Nelle immagini 5.11 e 5.12 si trovano i grafici della $g(x, y, 0)$ e della $g(x, 0, z)$ della fase nematica mostrata in figura 5.10, nella prima si mostra che sul piano ortogonale all'asse \hat{z} non sono presenti periodicità, questo esclude che sia una fase colonnare infatti il sistema ha un parametro d'ordine esagonale pari a $\psi_6 = 0.050$.

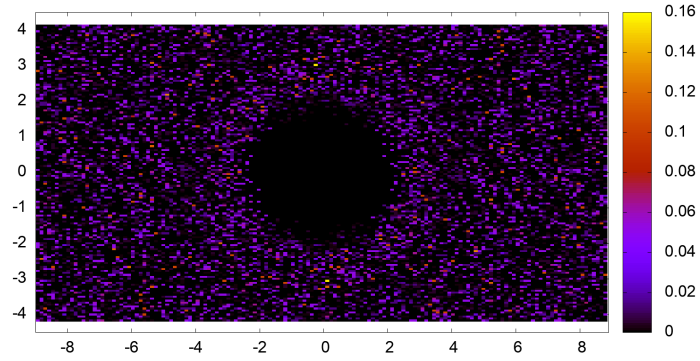


Figura 5.11. Funzione di distribuzione radiale $g(x, y, 0)$ con fase nematica e parametro d'ordine nematico $S = 0.764$. Il sistema non presenta periodicità sul piano $\hat{x} \hat{y}$.

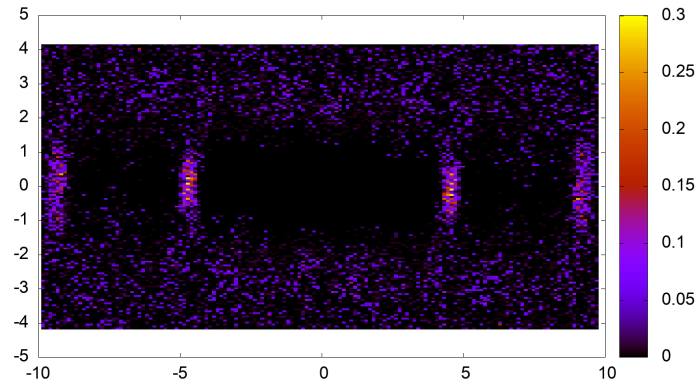


Figura 5.12. Funzione di distribuzione radiale $g(0, y, z)$ con fase nematica e parametro d'ordine nematico $S = 0.764$. Le molecole presentano una maggiore probabilità di avere delle molecole legate a se, mentre lateralmente non presenta periodicità.

Nella figura 5.12 si mostra che c'è una grande probabilità di trovare delle molecole legate che permettano la continuazione del filamento e che i filamenti non formano strutture a strati, visto che lateralmente alla molecola è presente un rumore senza periodicità. Andando a aumentare la pressione si vincolano ulteriormente le molecole a legarsi, inoltre si ottiene una fase nematica più pulita, questo è riscontrabile nel parametro d'ordine nematico che aumenta.

In conclusione, sono state trovate fasi nematiche che verranno utilizzate nel prossimo capitolo per la sostituzione con filamenti di semidischi. La compatibilità delle fasi nematiche con il paper pubblicato sul Langmuir [36] conferma inoltre che la limitazione nelle simulazioni a un numero di particelle ridotto ($N = 180$) non comporta dei problemi di taglia finita rilevanti.

5.4.2 Fase colonnare

Per verificare la stabilità della fase colonnare dopo la sostituzione con semidischi, si realizzano, analogamente a quanto fatto nel caso nematico, delle fasi colonnare con cilindri di *aspect-ratio* pari a due e quattro patch attrattive posizionate come nello scorso paragrafo. Si utilizzano $N = 180$ particelle e si lascia evolvere il sistema per 4 milioni di passi Monte Carlo partendo da una configurazione cristallina con il potenziale esterno dell'equazione 5.4 per i primi 10% dei passi. Variando la pressione si iniziano a ottenere fasi colonnari a partire da una frazione di volume pari a $\phi = 0.45$ e pressione ridotta pari a $P^* = 1.7$. Una fase colonnare, con parametro d'ordine esagonale $\psi_6 = 0.640$, è mostrata in figura 5.13.

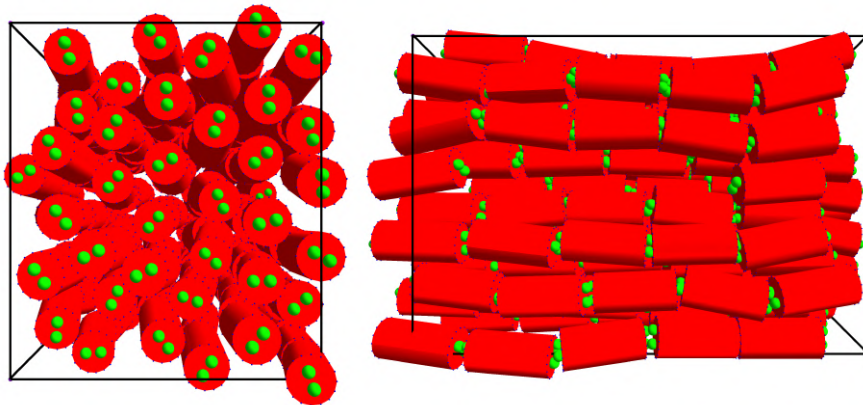


Figura 5.13. Vista frontale e laterale della fase colonnare con 180 molecole a una pressione ridotta di $P^* = 2$ e una temperatura ridotta di $T^* = 0.18475$. Con parametro d'ordine esagonale di $\psi_6 = 0.640$ e nematico di $S = 0.971$.

A differenza della fase nematica qui è presente un ordinamento esagonale sul piano $\hat{x} - \hat{y}$, per esplicitare questa periodicità si mostra il grafico della funzione di distribuzione radiale $g(x, y, 0)$ nella figura 5.14.

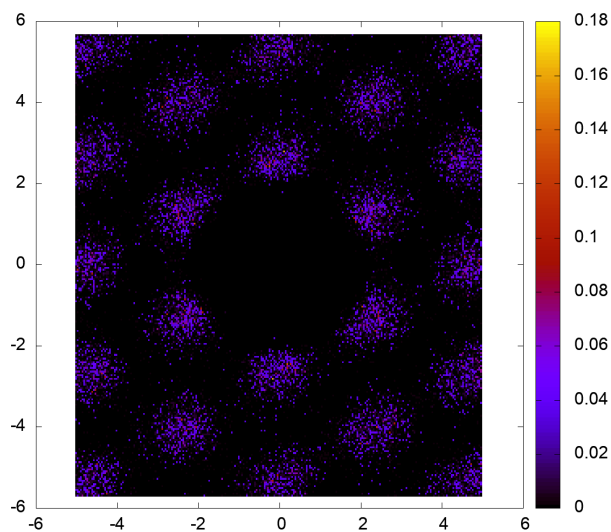


Figura 5.14. Funzione di distribuzione radiale $g(x, y, 0)$ della fase colonnare. Il sistema presenta una periodicità sul piano $\hat{x} \hat{y}$ esagonale.

Un'altra caratteristica della fase colonnare è l'assenza di una struttura a strati sul piano \hat{z} , questo la differenzia rispetto la fase smettica B mostrata nel paragrafo 1.4.2. L'assenza di una struttura a strati è mostrata dalla $g(x, 0, z)$ nell'immagine 5.15, dove si notano i due spot ad alta probabilità dove il filamento continua e, ai lati della molecola, delle fasce equiprobabili rispetto \hat{z} .

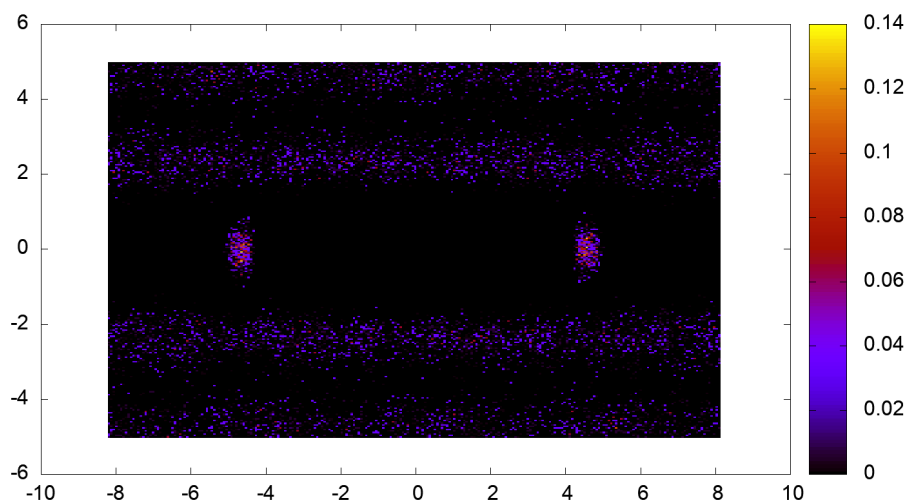


Figura 5.15. Funzione di distribuzione radiale $g(0, y, z)$ della fase colonnare. Le molecole presenta una maggiore probabilità di avere delle molecole legate a se così da formare il filamento, mentre lateralmente è presente la periodicità dovuta ai primi vicini della fase colonnare.

Andando a aumentare la pressione ridotta le fasi colonnari diventano più definite, questo è riscontrabile nell'aumento del parametro d'ordine esagonale.

5.4.3 Equazione di stato

Si simulano i cilindri a 4 patch attrattive così da poter costruire l'equazione di stato mostrando il cambiamento di fase tra isotropo-nematico e nematico-colonnare. Si utilizzano le simulazioni degli scorsi due paragrafi più altre simulazioni per costruire l'equazione di stato completa. Il risultato è mostrato in figura 5.16, dove si considerano fasi isotrope quelle con parametro d'ordine nematico $S < 0.15$ e si considerano fasi colonnari quelle con parametro d'ordine esagonale $\psi_6 > 0.3$.

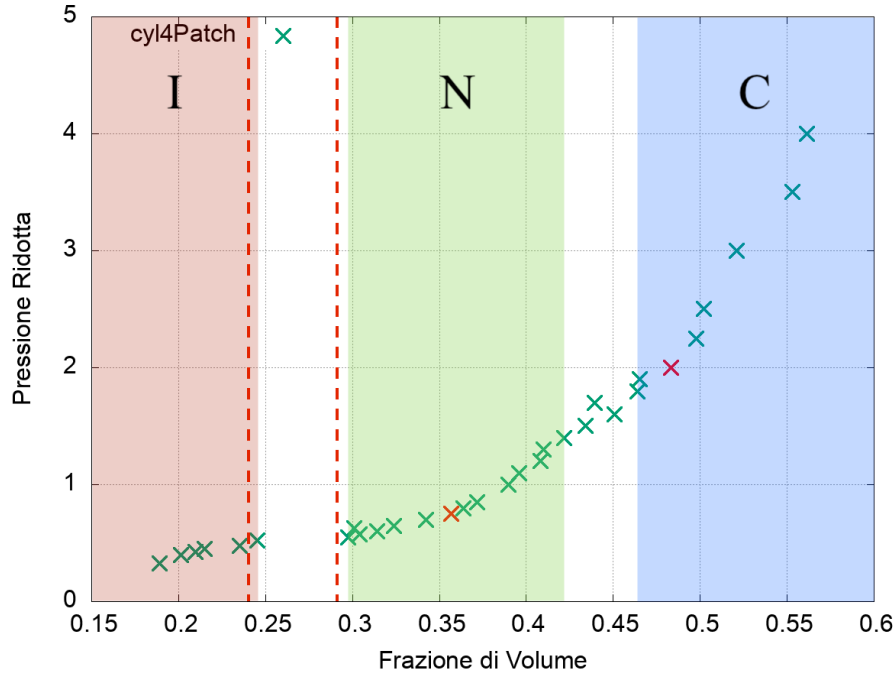


Figura 5.16. Equazione di stato per cilindri di *aspect-ratio* $X_0 = 2$ e 2 patch attrattive per base. Le simulazioni utilizzano $N = 180$ particelle e si evolvono per 4 milioni di passi Monte Carlo. I simboli in rosso saranno le due simulazioni utilizzate nelle sostituzioni con i semidischi, una nematica e una colonnare. Le rette tratteggiate in rosso rappresentano le transizione di fase isotropa-nematica a $T^* = 0.12$ per cilindri a due patch studiate nell'articolo [36], mostrate anche in figura 5.6.

Le imperfezioni nell'equazione di stato possono essere attribuite al numero di particelle limitato che si è utilizzato nelle simulazioni. Si noti inoltre che, mentre il cambiamento di fase isotropo-nematico è chiaro nelle simulazioni, quello nematico-colonnare è più incerto probabilmente poiché il sistema non ha completamente termalizzato. Nell'immagine 5.16 vengono anche indicate, con simboli rossi, le fasi che verranno utilizzate per le sostituzioni dei cilindri con semidischi. La fase colonnare viene presa più a ridosso della transizione di fase perché se stabile questa saranno stabili anche le successive a frazioni di volume maggiori.

Capitolo 6

Risultati

Si procede in questo capitolo a effettuare la sostituzione dei cilindri nella fase nematica e nella fase colonnare con filamenti composti da semidischi. Nel primo paragrafo si effettua la stima della temperatura ridotta per patch in posizione come nei semidischi. Inoltre si realizza una simulazione con un numero di semidischi ridotto per mostrare come i filamenti si formano a partire da una fase isotropa.

6.1 Energia libera di *stacking* per semidischi

Analogamente a quanto fatto nel paragrafo 5.3 si calcola la temperatura ridotta T_{2patch}^{*sd} tale che:

$$\Delta F_b^{1patch}(T_{1patch}^*) = \Delta F_b^{2patch}(T_{2patch}^{*sd}) \quad (6.1)$$

Nel caso di una patch per base si utilizza la stima già fatta e presente nella tabella in figura 5.4, mentre nel caso a due patch per base si posizionano quest'ultime sul cilindro come nei semidischi. Quindi, come descritto nel paragrafo 4.12.5, si posizionano patch a una distanza dal centro pari a $k = 0.35 \cdot r$ con un raggio pari a $r_{stacking} = 0.14$. I risultati ottenuti vengono presentati nella tabella in figura 6.1.

Numero di patch	Cilindro			Lato della box	Numero di sovrapposizioni		Inserimenti totali
	Altezza	Diametro	Volume		Di una patch	Di due patch	
2	4.2	2.1	14.1761	10	60578	\	5.00E+09
4	4.2	2.1	14.1761	10	13250	313	5.00E+09

Figura 6.1. Tabella con i risultati del programma per il calcolo dell'energia libera di *stacking*. Il volume dei cilindri viene calcolato come volume dei poliedri siccome vengono approssimati con 32 vertici. Nel caso a 4 patch totali il posizionamento è analogo a quello nei semidischi.

Si procede a calcolare l'equazione 5.1 nei due casi e a calcolare, interpolando, la temperatura ridotta T_{2patch}^{*sd} . Il processo è mostrato in figura 6.2, la temperatura ricavata è quindi $T_{2patch}^{*sd} = 0.13869$. Questa temperatura verrà utilizzata come indicazione per verificare in quale range di temperature avverrà lo scioglimento della fase una volta sostituiti i cilindri con i filamenti di semidischi.

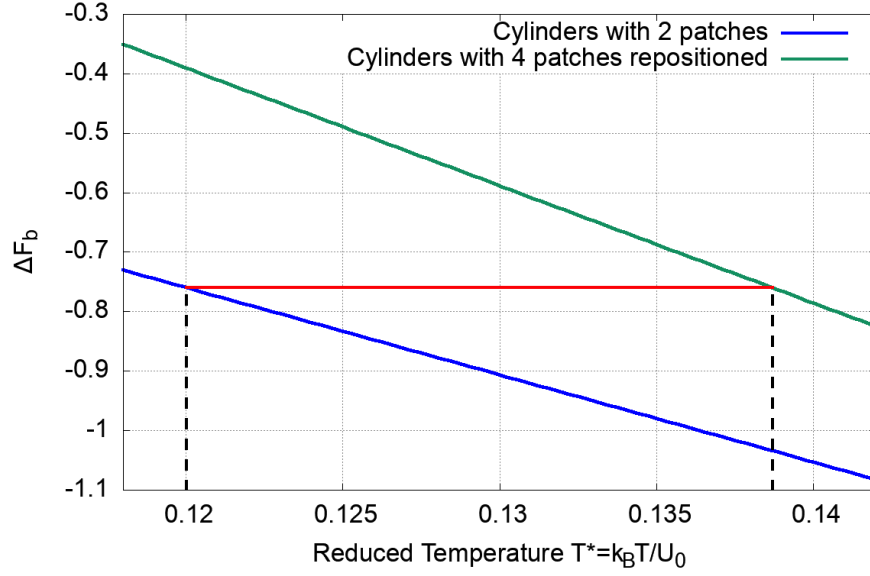


Figura 6.2. Andamento della ΔF_b in funzione della temperatura ridotta $T^* = k_B T / U_0$ per cilindri con 2 e 4 patch. Nel caso a 4 patch, queste vengono posizionate come nel caso dei semidischi, ovvero a $k = 0.35$ e con raggio $r_{stacking} = 0.14$.

6.2 Stabilità della fase nematica dopo la sostituzione con semidischi

Per verificare la stabilità delle fase nematiche, ottenute nel paragrafo 5.4, con la sostituzione dei cilindri con i filamenti di semidischi, si procede effettuando simulazioni NVT. La sostituzione con i filamenti va effettuata senza creare intersezioni tra gli oggetti, per questo il filamento dovrà essere tutto contenuto nel cilindro originale senza parti che sporgano al di fuori di esso. Questo avviene grazie al fatto che le simulazioni dei cilindri sono state effettuate con *aspect ratio* $X_0 = 2$ ma con raggio e altezza dei cilindri rispettivamente $r = 1.05$ e $h = 4.2$. Questo permette anche di inserire i due semidischi di $r = 1$ legati tramite *pairing* leggermente staccati, come visibile nell'immagine 6.3.

I cilindri e i semidischi vengono simulati con un numero finito di vertici, questo comporta che se si utilizza un numero di vertici maggiore per i semidischi questi potrebbero sporgere rispetto al cilindro iniziale. Quindi è necessario mantenere lo stesso livello di approssimazione tra i cilindri e i semidischi. Visto che le simulazioni dei cilindri sono state effettuate con un numero di vertici pari a $V = 32$, quindi 16 vertici per base, nel momento in cui si andrà a prendere la metà dei cilindri saranno necessari 9 vertici per base per avere lo stesso livello di approssimazione nei due oggetti. I semidischi verranno quindi simulati con un numero di vertici pari a $V = 18$, come visibile in figura 6.3.

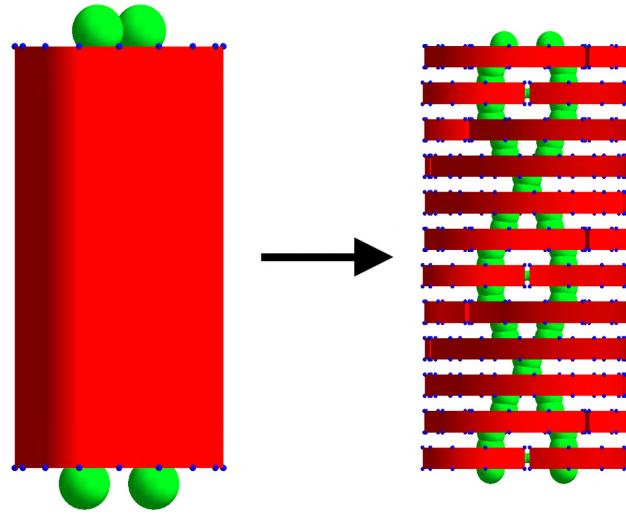


Figura 6.3. Cilindro composto da $V = 32$ vertici e filamento composto da 24 semidischi di $V = 18$ vertici.

La distanza tra le basi accoppiate del filamento viene scelta di modo che l'altezza finale sia la stessa del cilindro che viene sostituito. Il filamento viene anche costruito inserendo una rotazione di $\pi/5$ così da avere le patch di *stacking* già collegate tra loro. Per rappresentare il processo di sostituzione si prende la fase nematica a $P^* = 0.55$ e si mostra il processo di sostituzione in figura 6.4.

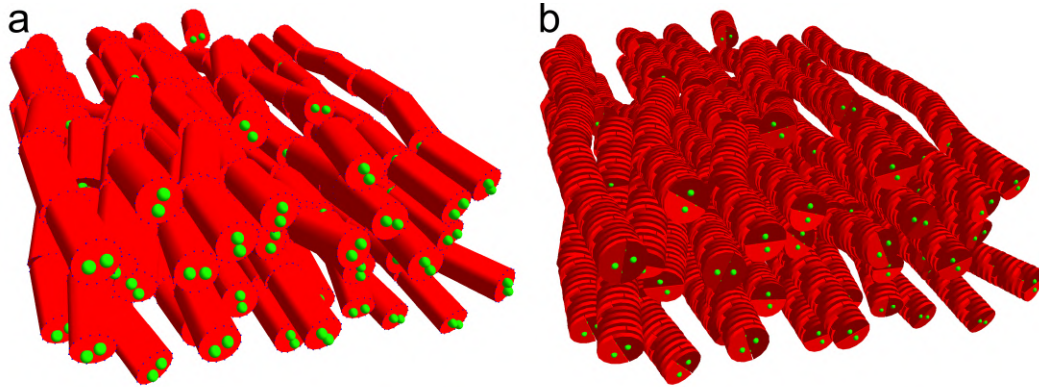


Figura 6.4. (a) Fase nematica con i cilindri a pressione $P^* = 0.55$. (b) Fase nematica dopo la sostituzione con i filamenti.

In figura 6.5 e 6.6 si mostrano rispettivamente la vista frontale e laterale della fase nematica, l'immagine viene suddivisa in due metà per mostrare il passaggio a filamenti di semidischi. La criticità della sostituzione è che le patch dei cilindri sono più grandi di quelle dei semidischi, questo comporta che solo alcuni legami tra cilindri verranno mantenuti. Tuttavia non è possibile diminuire di dimensione le patch dei cilindri, perché non aumenterebbe solo la loro *persistence length*, ma la diminuzione di flessibilità comporterebbe anche una impossibilità di ottenere una fase nematica.

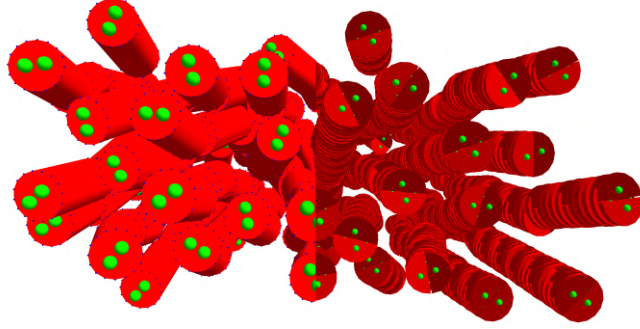


Figura 6.5. Vista frontale della fase nematica con sulla parte sinistra i cilindri e sulla destra la sostituzione con i semidischi.

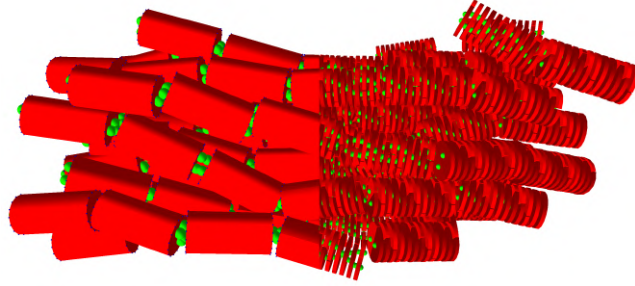


Figura 6.6. Vista laterale della fase nematica con sulla parte sinistra i cilindri e sulla destra la sostituzione con i semidischi.

Rimane da definire la profondità della buca delle patch di *pairing*, sperimentalmente il legame di *stacking* è circa di un ordine di grandezza più forte rispetto al legame di *pairing* come visto nel paragrafo 2.5, tuttavia questa differenza dipende da vari fattori come, ad esempio, dalla soluzione in cui si trovano i filamenti di DNA. Di conseguenza si opta in prima battuta di porre la stessa energia tra i legami di *pairing* e di *stacking*, perché se anche in questo caso si verifica lo scioglimento della fase allora avverrà sicuramente anche a legami di *pairing* meno forti. Per avere la stessa energia tra i legami di *stacking* e *pairing*, visto che i legami di *pairing* sono formati da due patch, si inserisce una profondità della buca di *pairing* dimezzata.

Si seleziona una fase nematica con frazione di volume $\phi = 0.356$ a pressione ridotta $P^* = 0.75$, si studia la stabilità con una simulazione nell'*ensemble* NVT. Si parte controllando la stabilità alla temperatura ottenuta nello scorso paragrafo $T_{2patch}^{*sd} = 0.13869$ per poi controllare se ci siano fasi stabili nell'intervallo di temperature $T = [0.1 : T_{2patch}^{*sd}]$. Si fa evolvere il sistema per almeno 3 milioni di passi Monte Carlo, tenendo monitorate l'energia totale del sistema e il parametro d'ordine nematico. Lo scioglimento della fase nematica avviene a tutte le temperature considerate infatti tutte le simulazioni presentano un parametro d'ordine nematico $S < 0.1$. Una simulazione in cui è avvenuto lo scioglimento è mostrata in figura 6.7. Il grafico riassuntivo, dove viene mostrato il parametro d'ordine nematico in funzione della temperatura, viene mostrato in figura 6.10.

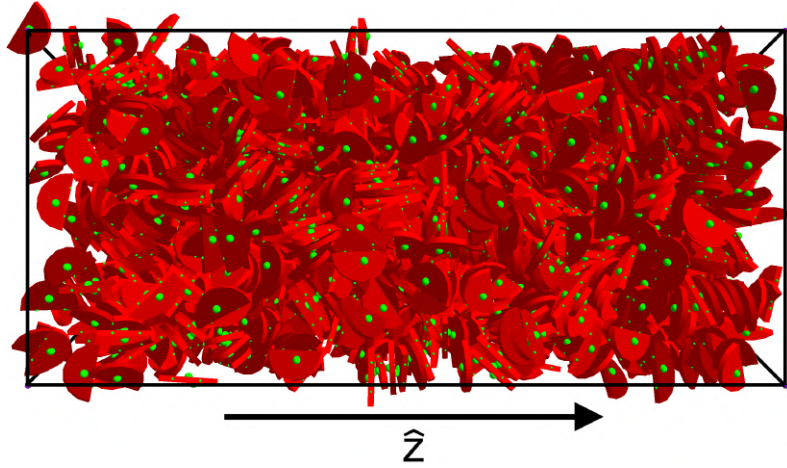


Figura 6.7. Fase senza ordinamento ricavata dallo scioglimento della fase nematica con $P^* = 0.5$. L'asse nematico della fase era inizialmente diretto parallelamente all'asse \hat{z} .

6.3 Stabilità della fase colonnare dopo la sostituzione con semidischi

In questo paragrafo si verifica la stabilità della fase colonnare una volta sostituiti i cilindri con filamenti di semidischi. Il processo è analogo a quanto descritto nella sostituzione con la fase nematica. Si seleziona quindi una fase colonnare dei cilindri a 4 patch attrattive con frazione di volume pari a $\phi = 0.480$ e pressione ridotta $P^* = 2$, questa è la prima fase colonnare con parametro d'ordine maggiore di 0.4. La fase colonnare scelta è mostrata in figura 6.8 con relativa sostituzione a semidischi e presenta un parametro d'ordine esagonale $\psi_6 = 0.640$.

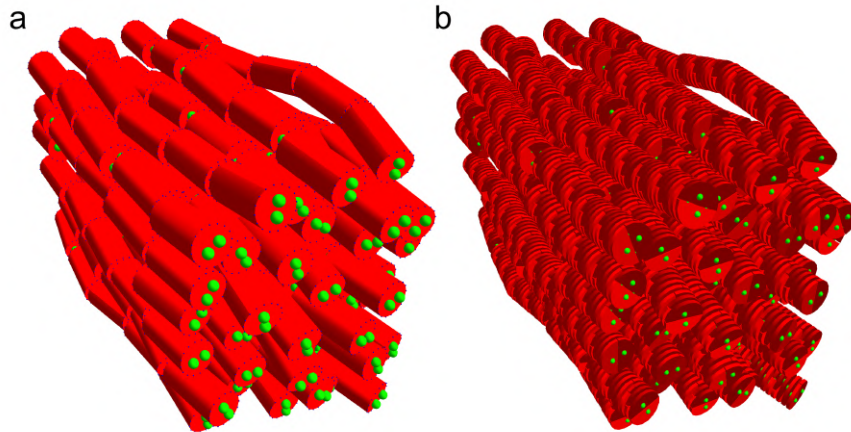


Figura 6.8. (a) Fase colonnare con i cilindri a pressione ridotta $P^* = 2$. (b) Fase colonnare dopo la sostituzione con i filamenti.

La funzione di distribuzione radiale $g(x, y, 0)$ di questa configurazione è mostrata in figura 5.14, mentre la vista frontale della fase è mostrata in figura 6.9.

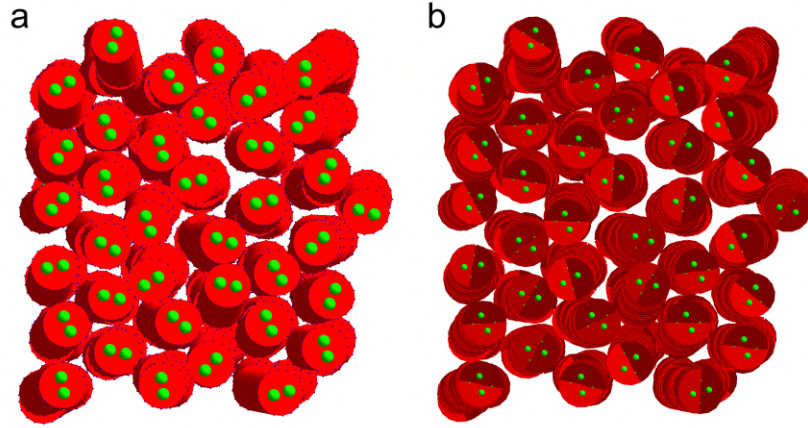


Figura 6.9. (a) Vista frontale della fase colonnare con i cilindri a pressione ridotta $P^* = 2$.
(b) Vista frontale della fase colonnare dopo la sostituzione con i filamenti.

Si effettua la sostituzione e si effettuano delle simulazioni nell'*ensemble* NVT nell'intervallo di temperature $T = [0.1 : T_{2patch}^{*sd}]$ dove $T_{2patch}^{*sd} = 0.13869$. Si lascia quindi evolvere il sistema per almeno 3 milioni di passi Monte Carlo.

Va precisato ora il metodo con cui viene calcolato il parametro d'ordine esagonale nel caso si simulino semidischi. In questo caso non si utilizza direttamente il centro del semidisco ma si controlla se questo sia legato a una seconda molecola a formare il disco e si considera il centro di quest'ultimo. Nel caso in cui non fosse legato si utilizza il centro del semidisco cosicché si ottenga, con una fase isotropa o nematica, un parametro d'ordine esagonale $\psi_6 \approx 0$ mentre solo nel caso di una fase colonnare perfetta con tutti i semidischi legati si ottenga $\psi_6 \approx 1$. Il parametro così modificato fornisce una indicazione sulla fase trovata, per avere la conferma si guarda alla funzione di distribuzione radiale $g(x, y, 0)$.

Nell'immagine 6.10 vengono mostrati i parametri d'ordine esagonali delle simulazioni partite dalla fase colonnare e i parametri d'ordine nematici dalle simulazioni partite dalle fasi nematiche ottenuti nello scorso paragrafo. Come detto nello scorso paragrafo le simulazioni che partivano dalla fase nematica mostrano lo scioglimento della fase. Mentre le simulazioni che partivano da fasi colonnari mostrano un parametro d'ordine esagonale non nullo tra le temperature $T^* = [0.1 : 0.11]$.

Per caratterizzare meglio le fasi a temperature inferiori a $T^* = 0.110$ si analizzano le funzioni di distribuzione radiale $g(x, y, 0)$. A temperature superiori a $T^* = 0.110$ si ottengono configurazioni parzialmente nematiche, con un parametro di ordine nematico $S \approx 0.5$ che si annulla per temperature superiori a $T^* = 0.13$. Abbassando la temperatura si iniziano a trovare fasi parzialmente colonnari, comprese tra le temperature ridotte $0.107 \leq T^* \leq 0.110$. In figura 6.11 (a) viene mostrata la simulazione equilibrata per $T^* = 0.110$. Per verificare che la simulazione abbia raggiunto l'equilibrio si guarda all'andamento del parametro d'ordine esagonale in funzione dei passi Monte Carlo, il caso di $T^* = 0.110$ è mostrato in figura 6.11 (b).

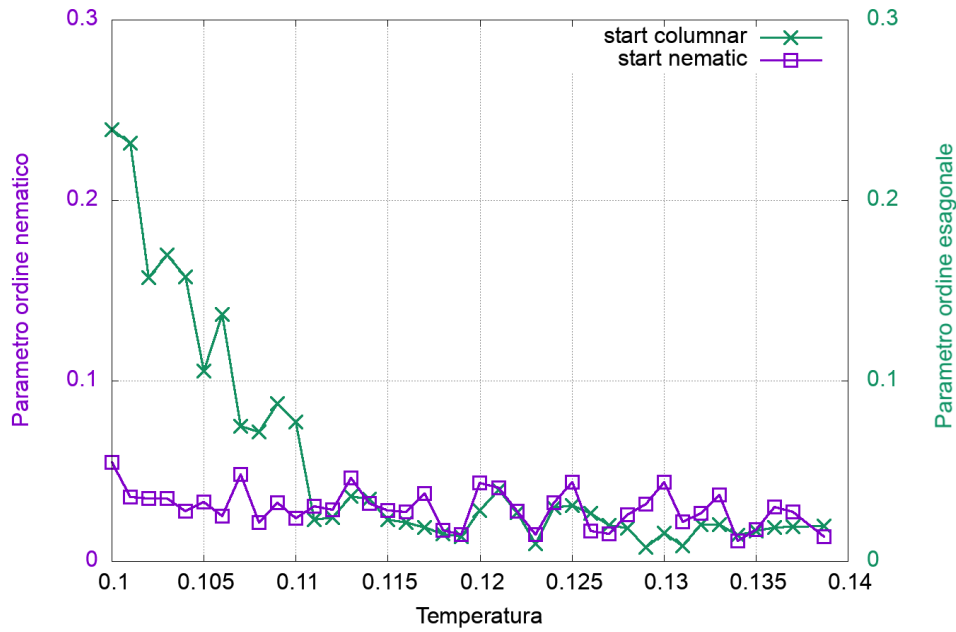


Figura 6.10. In verde il parametro d'ordine esagonale delle simulazioni che partivano da fasi colonnari in funzione della temperatura ridotta. In viola il parametro d'ordine nematico delle simulazioni che partivano da fasi nematiche in funzione della temperatura ridotta. Il grafico mostra la persistenza di fasi colonnari con semidischi e lo scioglimento delle fasi inizialmente nematiche.

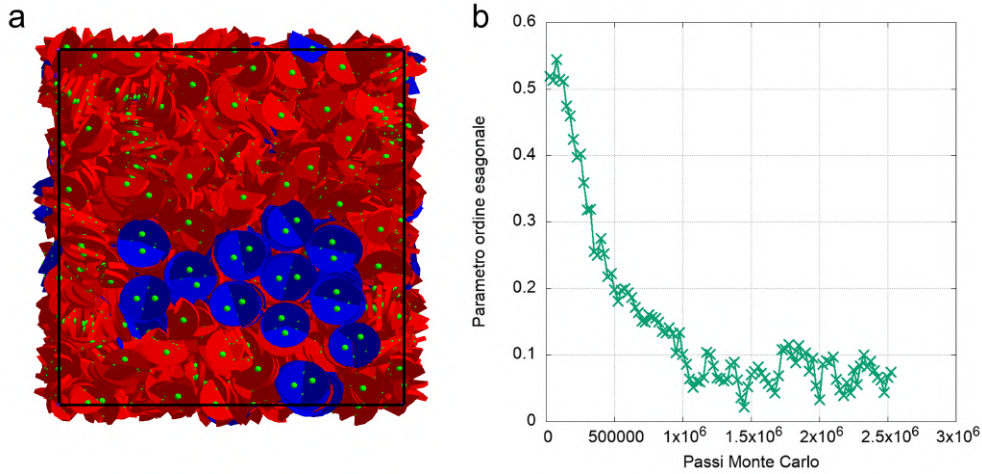


Figura 6.11. (a) Simulazione equilibrata a $T^* = 0.110$, in blu sono mostrati i semidischi che hanno entrambi i legami di *pairing* occupati. (b) Andamento del parametro d'ordine esagonale in funzione dei passi Monte Carlo a $T^* = 0.110$.

Le funzioni di distribuzione radiale $g(x, y, 0)$ per le temperature $T^* = 0.107$ e $T^* = 0.110$ sono mostrate in figura 6.12, come si nota c'è un ordinamento solo parziale delle molecole. Si specifica che la funzione di distribuzione radiale viene realizzata guardando al centro di ogni semidisco e non al centro del disco eventualmente formato.

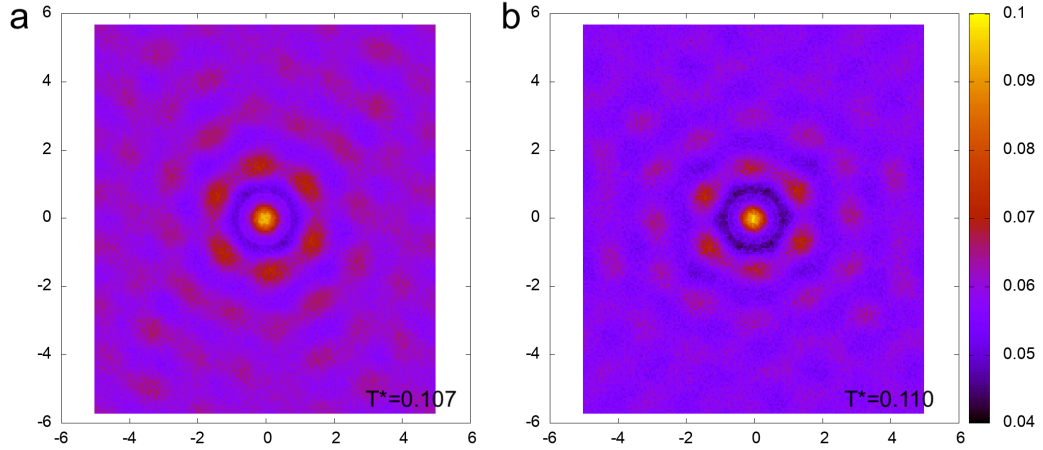


Figura 6.12. (a) Funzione di distribuzione radiale $g(x, y, 0)$ a temperatura $T^* = 0.107$. (b) Funzione di distribuzione radiale $g(x, y, 0)$ a temperatura $T^* = 0.110$.

Andando a abbassare ulteriormente la temperatura, per $T^* \leq 0.106$, si ottengono delle fasi colonnari più chiare, questo è riscontrabile anche vedendo il numero dei filamenti che rimangono completi, basti guardare alle configurazioni a $T^* = 0.100$ e $T^* = 0.104$ mostrate in figura 6.13.

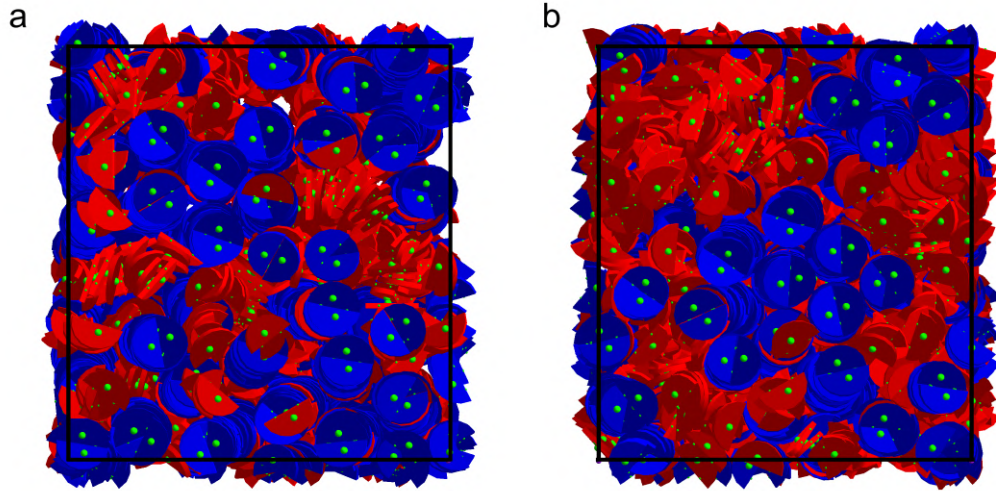


Figura 6.13. (a) Simulazione equilibrata a $T^* = 0.100$, in blu sono mostrati i semidischi che hanno entrambi i legami di *pairing* occupati. (b) Simulazione equilibrata a $T^* = 0.104$.

Queste simulazioni presentano un parametro d'ordine esagonale più grande e, andando a guardare la $g(x, y, 0)$, mostrata per le temperature $T^* = 0.100$ e $T^* = 0.104$ in figura 6.14, è riscontrabile una periodicità molto più netta.

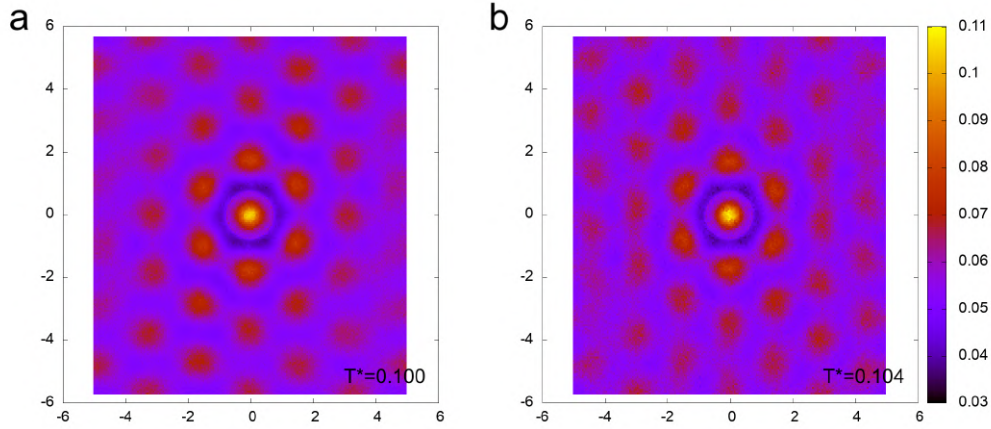


Figura 6.14. (a) Funzione di distribuzione radiale $g(x, y, 0)$ a temperatura $T^* = 0.100$. (b) Funzione di distribuzione radiale $g(x, y, 0)$ a temperatura $T^* = 0.104$.

Ricapitolando si è visto che la fase nematica, una volta sostituiti i cilindri con i semidischi, nell'intervallo di temperature $T^* = [0.1 : 0.14]$ non è stabile, mentre la fase colonnare presenta una zona tra le temperature $T^* = [0.1 : 0.106]$ dove la periodicità rimane stabile. Rimane da discutere come avviene lo scioglimento della fase colonnare, se i filamenti vengono mantenuti o se questi vengono separati nei singoli semidischi. Si procede analizzando l'energia, proporzionale al numero di legami presenti, raggiunta dalle simulazioni dopo 2.5 milioni di passi Monte Carlo al variare della temperatura ridotta, il grafico è mostrato in figura 6.15.

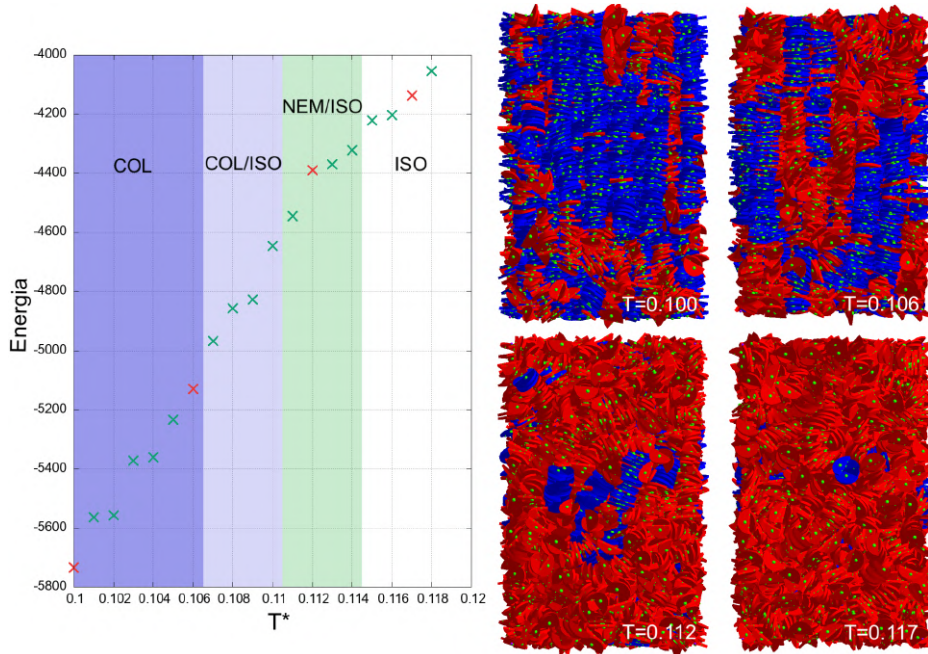


Figura 6.15. Grafico dell'energia in funzione di T^* dopo 2.5 milioni di passi Monte Carlo. La fase isotropa-nematica presenta un parametro d'ordine nematico compreso tra $0.4 < S < 0.2$. Le simulazioni indicate con il simbolo rosso nel grafico sono mostrate a lato, in blu si indicano i semidischi che hanno entrambe le patch di *pairing* occupate.

Si nota che il numero di legami diminuisce in funzione della temperatura e che i legami di *pairing* vengono quasi totalmente rimossi a alte temperature. Infatti per $T^* = 0.117$ si ha in media circa un legame per semidisco e la quasi totalità sono legami di *stacking*, come mostrato in figura 6.15, quindi nella fase isotropa sono presenti dei filamenti formati da semidischi legati solo tramite *stacking*. Tuttavia questi filamenti non si allineano e quindi non si forma la fase nematica a alte temperature.

6.4 Autoassemblamento dei filamenti

Per verificare che la costruzione dei filamenti avvenga correttamente si realizza una simulazione in *ensemble* NVT con un numero di semidischi ridotto $N = 800$ a temperatura ridotta $T^* = 0.10$ e frazione di volume $\phi = 0.1610$. Il processo di costruzione dei filamenti viene mostrato in figura 6.16 a variare del numero di passi Monte Carlo.

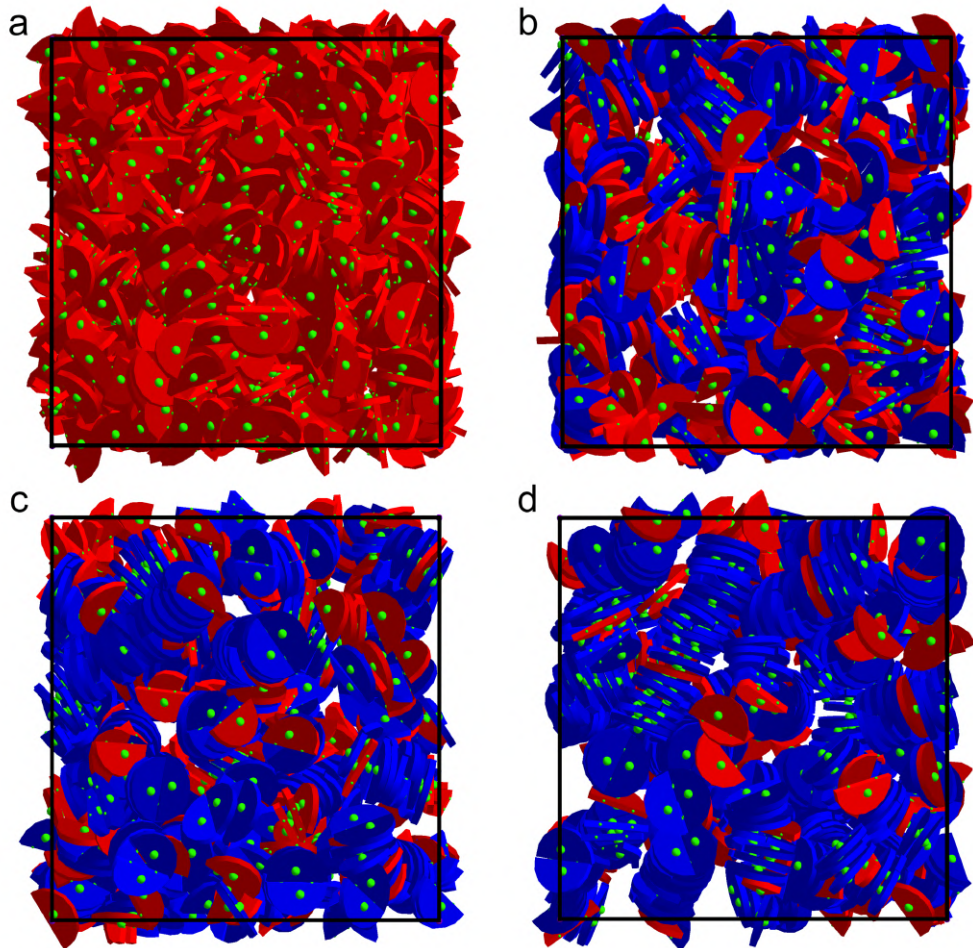


Figura 6.16. (a) Configurazione di partenza della simulazione. Configurazione della simulazione dopo due (b), quattro (c) e sei (d) milioni di passi Monte Carlo. In blu sono mostrati i semidischi che hanno entrambi i legami di *pairing* occupati.

Nell'immagine 6.17 viene mostrata la simulazione dopo otto milioni di passi Monte Carlo, vengono inoltre evidenziati per il filamento più lungo le separazioni tra semidischi che formano un disco per rendere più chiara la rotazione del filamento. La simulazione mostra come i filamenti si formino e, grazie al posizionamento delle patch di *stacking*, come questo ruoti su se stesso.

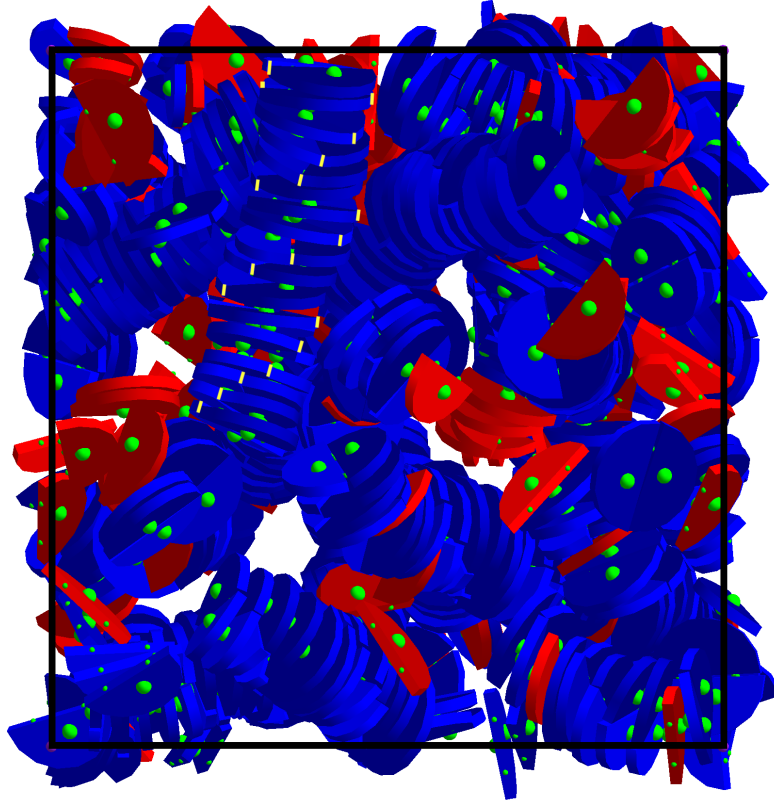


Figura 6.17. Configurazione della simulazione dopo otto milioni di passi Monte Carlo. Vengono evidenziate le separazioni del filamento più lungo per rendere chiara la rotazione del filamento.

6.5 Confronto con dati sperimentali

Si vogliono confrontare i risultati ottenuti con i filamenti di semidischi con dei risultati sperimentali, per verificare se ci sia un accordo tra le stime ottenute. I risultati sperimentali vengono presentati utilizzando la concentrazione in unità di misura (mg/ml), di conseguenza i risultati ottenuti dalle simulazioni andranno ricalcolati utilizzando l'equazione 6.2.

$$c = \frac{N}{V} m_N \quad (6.2)$$

Dove c è la concentrazione, N è il numero di semidischi contenuti nel volume V della box di simulazione e m_N è la massa molecolare di un singolo nucleotide.

Questa massa, misurata in Dalton, è pari a $m_N = 482.2 Da$ per filamenti di dTTP e $m_N = 491.2 Da$ per filamenti di dATP. I risultati sperimentali a disposizione sono per uno studio di una soluzione di filamenti dTTP+dATP, di conseguenza si utilizza una media tra le due masse pari a $m_N = 486.7 Da$. Utilizzando l'equazione 6.2 si ottiene:

$$\begin{aligned} c_{NEM} &= 231.1 (Da/nm^3) = 383.8 (mg/ml) \\ c_{COL} &= 398.0 (Da/nm^3) = 660.8 (mg/ml) \end{aligned} \quad (6.3)$$

Con c_{NEM} si indica la concentrazione della simulazione che partiva dalla fase nematica dei cilindri ma che diventa isotropa nel momento in cui si sostituiscono i semidischi. Analogamente c_{COL} indica la concentrazione della simulazione che partiva dalla fase colonnare e che presenta un range di temperature che mantiene tale fase anche con i semidischi. Si confrontano le concentrazioni ottenute con i risultati presenti nella Ref. [53].

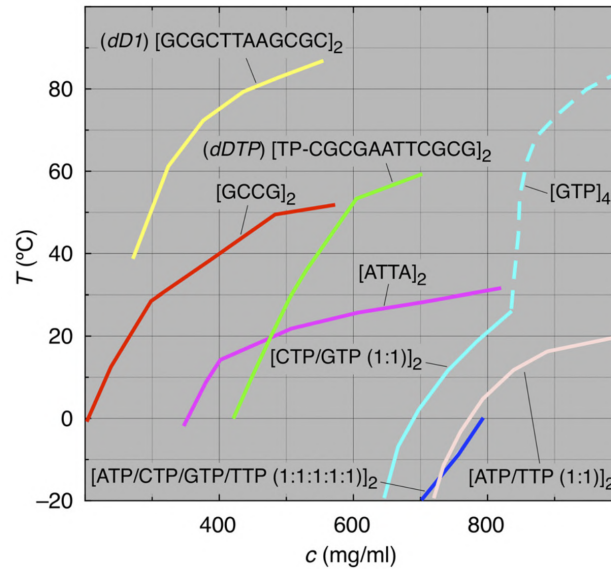


Figura 6.18. Raccolta dei relativi diagrammi di fase LC del DNA di oligomeri e monomeri, le linee indicano la temperatura massima a cui avviene l'ordinamento liquido cristallino. La transizione di fase si sposta a più basse concentrazioni e più alte temperature con l'aumento della lunghezza dei oligomeri, indicando che la stabilità delle fasi LC è aumentata dalla polimerizzazione. Immagine tratta da [53].

Nell'articolo vengono analizzate le fasi liquido cristalline di diversi tipi di filamenti di DNA, che variano per struttura e sequenza nucleotidica. La concentrazione e la temperatura a cui avviene il cambiamento di fase varia per ogni tipo di filamento, come mostrato nella figura 6.18. Gli esperimenti, discussi nella Ref. [53], vengono realizzati utilizzando nucleotidi trifosfato (dNTPs) acquistati già in soluzione. La preparazione del campione è mostrata in figura 6.19, il dNTP viene inizialmente essiccato tramite un processo di liofilizzazione e successivamente viene bloccata la possibile evaporazione nei successivi step coprendo il campione con un olio minerale. Viene quindi reidratato il campione inserendo una quantità predefinita di acqua per raggiungere la concentrazione desiderata.

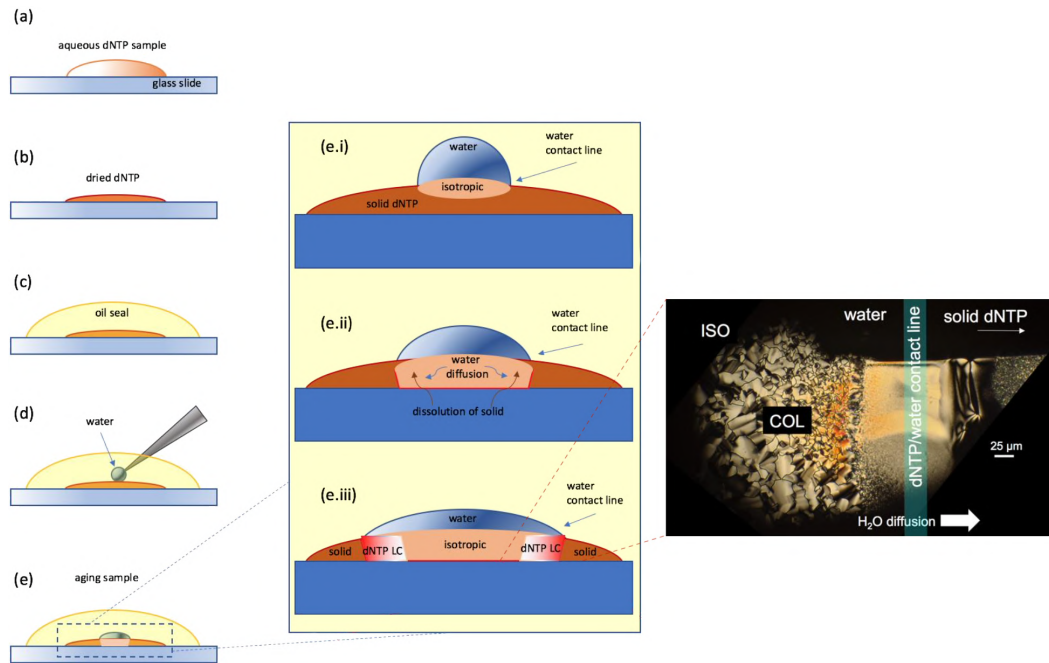


Figura 6.19. (a) Si posiziona la soluzione di dNTP sul vetrino. (b) Si utilizza il vuoto per liofilizzare il campione. (c) Si posiziona una goccia di olio minerale per minimizzare l'evaporazione. (d) Si reidrata il campione. (e) Si lascia riposare il campione a temperature vicine alla transizione ISO/COL. (e.i) Si formano fase isotrope tra l'interfaccia solido liquida. (e.ii) L'acqua si propaga nel campione allargando la zona isotropa. (e.iii) La diffusione dell'acqua rallenta e si formano fasi liquido cristalline sull'interfaccia, visibili nell'immagine a PTOM inserita per soluzioni dATP/dCTP/dGTP/dTTP (1:1:1:1). Immagine tratta dalle *Supplementary Information* della Ref. [53].

Successivamente il campione viene analizzato attraverso la diffrazione a raggi X (XRD) e la microscopia ottica a trasmissione polarizzata (PTOM). La XRD è una tecnica utilizzata per determinare la struttura cristallina di un materiale. Si irradia un campione del materiale con raggi X incidenti e analizzando l'intensità dei raggi X e la loro posizione in funzione dell'angolo di diffusione è possibile determinare la struttura del materiale. La microscopia ottica a trasmissione polarizzata è stata discussa nel paragrafo 1.2.1 e lo schema di funzionamento è stato mostrato in figura 1.1 nel sopracitato paragrafo.

In particolare nell'articolo si analizza il diagramma di fase degli dATP/dTTP, mostrati in figura 6.20. In questo caso si ha una transizione di fase isotropo-colonnare come trovato nelle simulazioni con i semidischi. In figura 6.21 si confrontano i risultati ottenuti sperimentalmente per i filamenti di dATP/dTTP [53] con i risultati ottenuti in questa tesi tramite simulazioni Monte Carlo di semidischi.

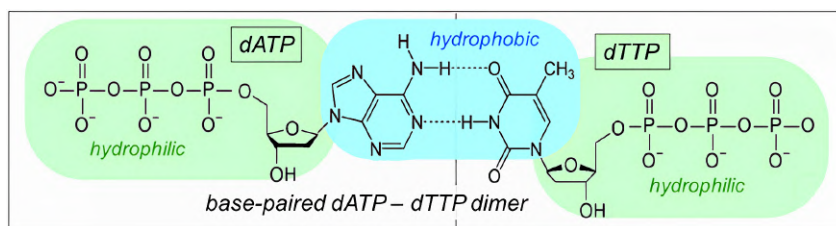


Figura 6.20. Struttura chimica del dimero dATP/dTTP con indicazione dei domini idrofobi e idrofili. Immagine tratta della Ref. [53].

Le simulazioni che partono da concentrazioni più alte mostrano un cambio di fase per $T^* = 0.106$, mentre nel caso a concentrazione minore questo cambio di fase non è presente. Questo risultato è compatibile con l'andamento sperimentale, come mostrato nella figura 6.21. Alzando la temperatura nelle simulazioni si passa dalle fasi isotrope-colonnari a fasi isotropo-nematiche, quest'ultime potrebbero essere dovute alla non equilibratura del sistema che, a causa dell'alto numero di particelle e all'alta concentrazione, richiede molto tempo o potrebbero essere dovute a problemi di taglia finita.

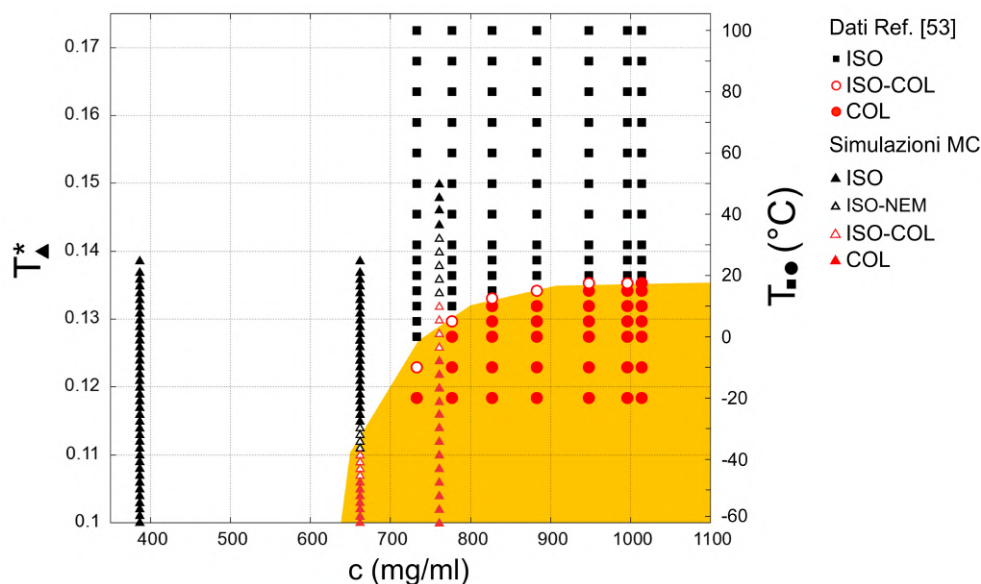


Figura 6.21. Diagramma di fase dei filamenti di DNA dTTP/dATP. I simboli quadrati e tondi fanno riferimento ai risultati sperimentali mentre i simboli triangolari fanno riferimento ai risultati ottenuti dalle simulazioni. In arancione viene indicata la zona dove è presente la fase colonnare. I dati sperimentali sono stati resi disponibili dal professor Bellini.

Successivamente è possibile confrontare i risultati delle simulazioni con gli esperimenti con filamenti dCTP+dGTP. In questo caso la massa molecolare del dCTP è $m_N = 467.2 Da$, mentre del dGTP è $m_N = 507.2 Da$, analogamente a prima si utilizza la media dei due valori: $m_N = 487.2 Da$.

Utilizzando l'equazione 6.2 si ottiene:

$$\begin{aligned} c_{NEM} &= 231.3 \text{ (Da/nm}^3\text{)} = 384.1 \text{ (mg/ml)} \\ c_{COL} &= 398.4 \text{ (Da/nm}^3\text{)} = 661.6 \text{ (mg/ml)} \end{aligned} \quad (6.4)$$

Si confronta nella figura 6.22 il diagramma di fase con i risultati presenti nella Ref. [53] per filamenti dCTP+dGTP con i risultati ottenuti dalle simulazioni Monte Carlo utilizzando semidischi. In questo caso l'andamento sperimentale mostra un ulteriore salto in temperatura a alte concentrazioni dovuto alla formazione di fasi colonnari formate da quadruplex, ovvero strutture di DNA a quadrupla elica. Questo cambio di struttura non è permesso nelle simulazioni per come sono state posizionate le patch di conseguenza il modello non potrà seguire l'andamento sperimentale a alte concentrazioni.

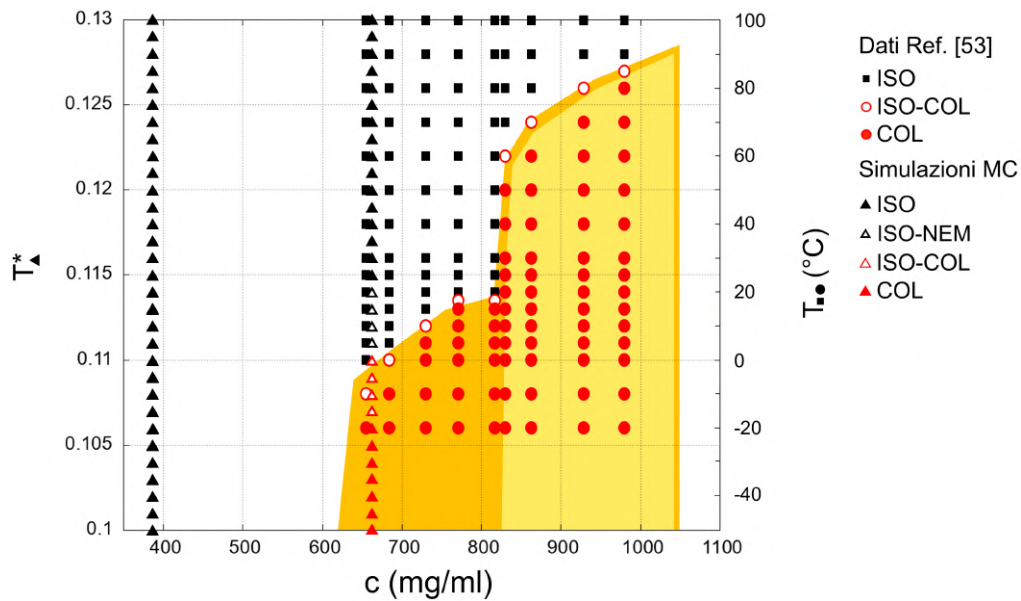


Figura 6.22. Diagramma di fase dei filamenti di DNA dCTP/dGTP. I simboli quadrati e tondi fanno riferimento ai risultati sperimentali mentre i simboli triangolari fanno riferimento ai risultati ottenuti dalle simulazioni. In arancio viene indicata la zona dove è presente la fase colonnare, mentre la zona gialla mostra la zona colonnare formata da DNA a quadrupla elica (quadruplex). I dati sperimentali sono stati resi disponibili dal professor Bellini.

Le fasi previste dalle simulazioni sono in accordo con i dati sperimentali ma, un possibile sviluppo futuro che confermerebbe questa tesi, potrebbe essere quello di effettuare simulazioni a concentrazioni più alte così da poter confrontare l'intero diagramma di fase con quello sperimentale sia per le soluzioni dCTP/dGTP che dTTP/dATP.

Capitolo 7

Conclusioni

In questo capitolo si esporrà una sintesi esaustiva dei risultati ottenuti nella tesi, andando a analizzare i punti di forza e le problematiche del modello realizzato. Si analizzeranno gli algoritmi utilizzati e il vantaggio portato dalla loro ottimizzazione, la possibilità di generalizzare il programma scritto per essere utilizzato con tutti gli oggetti convessi e non convessi. Si presenterà brevemente come il modello a semidischi è stato modificato e regolato per adattarsi alla descrizione di singoli nucleotidi. Infine si discuteranno i risultati ottenuti dal modello e le prospettive che questi studi sul DNA hanno da offrire in ambito scientifico. I punti fondamentali dei risultati ottenuti in questa tesi sono i seguenti:

- Nella tesi si sono implementati due algoritmi per l'intersezione di poliedri convessi, il *SAT* e *Xenocollide*, questo ha permesso di confrontarli per aver sicurezza che i risultati combacino. La scelta è ricaduta su *Xenocollide*, visto che risulta circa quattro volte più veloce rispetto al *SAT* per il numero di vertici utilizzato nella tesi. Sono state poi utilizzate le *Linked cell list* con due liste separate, una per le molecole simulate e un'altra per le patch attrattive di queste molecole.
- Il vantaggio nell'approssimazione degli oggetti in poliedri con un numero di vertici finito risulta nella semplicità con cui è possibile cambiare forma e implementare tagli nelle molecole. Gli algoritmi disponibili in letteratura permettono di trovare intersezioni di oggetti regolari e la loro ottimizzazione è strettamente legata alla regolarità di questi oggetti. Qui il limite risulta solo nella simulazione di oggetti convessi, limite che però potrebbe essere superato implementando nel programma un metodo di scomposizione di oggetti concavi in un numero finito di oggetti convessi, così da poter continuare a utilizzare *Xenocollide*. La criticità di questo metodo risulta nel numero di vertici con cui si approssimano gli oggetti, per questo nella tesi si è confrontata l'equazione di stato dei cilindri (non approssimati) con quella dei cilindri approssimati con $V = 16, 24, 32$ vertici. Verificando quindi che con $V = 32$ vertici si aveva già un buon accordo tra le due equazioni di stato.
- Si è studiata quindi l'equazione di stato dei cilindri a 4 patch attrattive verificando la presenza delle tre fasi, isotropa, nematica e colonnare. La dimensione e la posizione delle patch sono state scelte per essere nelle stesse

posizioni nel momento in cui avverrà la sostituzioni con i semidischi e per avere una flessibilità dei filamenti costruiti paragonabile a quella sperimentale dei filamenti di DNA.

- Si è adattato il modello di semidischi per avere parametri simili ai filamenti di DNA. In particolare si è ottenuta una *persistence length* compatibile con quella del DNA e una lunghezza dei filamenti costruiti simile a quella sperimentale. Inoltre le patch di *stacking* sono state ruotate per costruire filamenti che si avvolgono su se stessi per avere la stessa rotazione dei filamenti di DNA; la rotazione permette di avere una flessibilità che non dipenda dalla direzione rispetto al taglio dei semidischi.
- Infine sono state prese una fase nematica e una colonnare, ottenute con cilindri a 4 patch, e sono stati sostituiti i cilindri di *aspect-ratio* $X_0 = 2$ con filamenti composti da 24 semidischi. Si studia la stabilità di queste fasi al variare della temperatura. Si nota che è presente un range di temperatura dove la fase colonnare risulta stabile a differenza della fase nematica. Questo risultato si differenzia dagli altri modelli *coarse-grained* utilizzati in letteratura dove entrambe le fasi risultavano stabili alla stessa temperatura. La scelta di partire da fasi già definite è giustificata dal fatto che, partendo da una configurazione isotropa con questo numero di semidischi ($N = 4320$), i tempi di equilibratura sarebbero stati troppo lunghi.
- Si sono confrontati i risultati ottenuti con i valori sperimentali presenti in letteratura per soluzioni di filamenti dCTP/dGTP e dTTP/dATP, mostrando che il modello presenta fasi LC compatibili con quelle sperimentali. L'interesse per le fasi LC del DNA e dell'RNA deriva da alcune ricerche recenti in cui si discute la possibilità che queste fasi abbiano facilitato la formazione di filamenti in uno stato primordiale della vita, avendo quindi un ruolo nell'origine di essa [23].

In conclusione il modello, pur essendo più costoso in termini di tempo computazionale rispetto al modello con i cilindri, permette di avere temperature con fasi colonnari stabili e fasi nematiche non stabili in accordo con quanto trovato sperimentalmente nella Ref. [53]. Un obiettivo ulteriore potrebbe essere quello di ottimizzare il programma, cercando algoritmi di intersezione di poliedri più veloci grazie ai repentini sviluppi in questo ambito dovuti all'interesse nelle simulazioni 3D, e riuscire a trovare fasi liquido cristalline partendo da configurazioni iniziali isotrope con delle simulazioni in *ensemble* NPT.

Successivamente si potrebbe utilizzare il programma per effettuare simulazioni di altri oggetti di interesse scientifico, confrontando i risultati ottenuti al variare del numero di vertici con cui avviene l'approssimazione con i risultati presenti in letteratura. Anche se gli algoritmi di intersezione di oggetti regolari, come cilindri o ellissoidi, risultano più veloci per il grado di precisione che offrono, visto che non si approssimano a poliedri, questo metodo permette una flessibilità maggiore nella scelta della forma. Permette di avere oggetti non regolari con tagli e deformazioni, per ora solo con oggetti convessi, purché si utilizzi un numero di vertici sufficiente nell'approssimazione.

Appendice A

Implementazione e ottimizzazione codice

A.1 Patch con esclusività di legame

Si presentano ora diverse impostazioni che è possibile attivare nel programma realizzato in C++ riguardanti l'interazione tra patch. La base dell'algoritmo è quella di costruire un array `INT` lungo quanto il numero di patch totali, quindi pari al numero di particelle totali moltiplicato il numero di patch per particella. Si consideri ora che la particella i -esima interagisca con la particella j -esima allora questa intersezione verrà salvata con $\text{INT}[i] = j$ e con $\text{INT}[j] = i$. Questo permette di controllare velocemente se una patch è già occupata da un legame ma anche di poter eliminare un vecchio legame, infatti se si sposta la particella della patch i si rimuove il legame come $\text{INT}[\text{INT}[i]] = -1$ e $\text{INT}[i] = -1$.

- **Singola interazione:** Questa funzione limita a un legame ogni patch, non permettendo quindi doppi legami nel caso la conformazione della particella lo permetta. Ad esempio se si aumenta il raggio della patch di *pairing* c'è la possibilità che questa si leghi più di una volta. Per limitare a un legame è sufficiente accettare la mossa solo nel caso in cui la patch i e j rispettino che $\text{INT}[i] = -1 \&\& \text{INT}[j] = -1$.
- **Interazione selettiva:** Si imposta a inizio programma il tipo delle patch bloccando l'interazione a solo patch dello stesso tipo, così da poter escludere il legame tra patch di *stacking* e di *pairing*. Si realizza rifiutando le mosse con patch interagenti di tipo diverso.
- **Interazione tra patch opposte bloccata:** Nel caso in cui le patch di *stacking* abbiano un raggio superiore alla metà dell'altezza del semidisco è possibile che si crei un doppio legame di stacking, ovvero che le due patch di *stacking* esterne di un filamento composto da 2 semidischi interagiscano con le interne di semidischi opposti, come mostrato in figura A.1. Si verrebbe così a creare un legame a energia doppia e, se l'opzione **singola interazione** fosse attiva, a bloccare l'ulteriore costruzione del filamento. Per risolvere il problema, nel caso in cui $2r_{stacking}^{patch} > h_{semidisco}$, si attiva

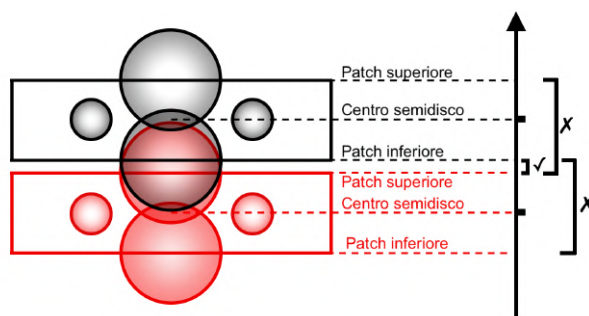


Figura A.1. Rappresentazione schematica dell'interazione tra patch di *stacking* su basi opposte. La patch superiore del semidisco sopra tocca la patch superiore del semidisco sotto e viceversa. L'unica interazione valida è solo quella tra le due patch interne.

automaticamente un controllo. Il programma verifica che entrambe le proiezioni, sulla retta che unisce i centri dei due semidischi, dei centri delle patch interagenti caschino tra i centri dei due semidischi, altrimenti, anche se solo una delle due patch è fuori, le considera come trasparenti tra loro. Il procedimento è mostrato schematicamente in figura A.1.

- **Pairing esclusivo:** Questa funzione impone che le patch di *pairing* di una particella siano legate entrambe con la stessa particella, in altre parole se la prima patch di *pairing* della particella i è legata con la particella j allora la seconda patch di *pairing* della particella i potrà legarsi solo con la seconda patch della particella j .
- **Blocco delle interazioni:** Questa funzione mantiene i legami delle patch della configurazione iniziale, è stata utilizzata per stimare la *Persistence length* di un filamento composto da dischetti tagliati. Infatti per capire la flessibilità del filamento era necessario che i legami non si rompessero durante le mosse di traslazione e rotazione.

A.2 Linked cell list

Per determinare l'energia del sistema o per valutare se ci sono sovrapposizioni tra sfere dure è necessario valutare la distanza per ogni coppia di particelle \mathbf{r}_{ij} , verificare che questa non sia minore della distanza minima di interazione e, eventualmente, calcolare il potenziale. Tuttavia, se si utilizza questo metodo, il tempo per completare la simulazione va come N^2 dove N è il numero delle particelle. Per ottimizzare questo andamento è possibile utilizzare l'algoritmo delle *Linked cell list*.

Per utilizzare l'algoritmo delle *Linked cell list* si divide il box dove avviene la simulazione (in genere un volume cubico di lato L) in un numero NCELL di celle uguali, queste celle hanno la prerogativa di avere una dimensione maggiore di un eventuale raggio di cutoff r_c dell'interazione. Nel caso l'interesse sia solo verificare la sovrapposizione di poliedri il lato della cella dovrà essere almeno pari alla distanza tra i due vertici più lontani della stessa molecola. Una particella, che si trova in una certa cella, potrà interagire solo con particelle che si trovino entro le 26 (8 nel caso bidimensionale) celle adiacenti alla propria e quelle contenute nella sua

stessa cella. Di conseguenza, se due particelle hanno una distanza maggiore della lunghezza della cella, queste non possono interagire così da poter risparmiare il tempo computazionale che ora avrà un andamento come N .

La struttura delle celle viene definita mediante il metodo delle *linked lists*. Si inizia collocando ogni particella nella cella corretta. Durante questo processo si creano due arrays. L'array **HEAD**, di dimensione pari a quella del numero delle celle, contiene il numero di identificazione di una particella contenuta nella cella *icell* (dove *icell* è il numero che identifica la cella). Questo numero punta a un elemento dell'array **LIST**, di lunghezza N , che contiene l'indice della particella successiva in quella cella.

Dunque un elemento di **LIST** contiene l'indice di una certa particella (ad esempio i); questo numero punta a sua volta all' i -esimo elemento di **LIST**, che contiene il numero di un'altra particella e così via. Il procedimento continua finché non si raggiunge l'elemento di **LIST** che contiene il valore -1 , ovvero il valore di stop. Il processo, nel caso bidimensionale è mostrato nella figura A.2.

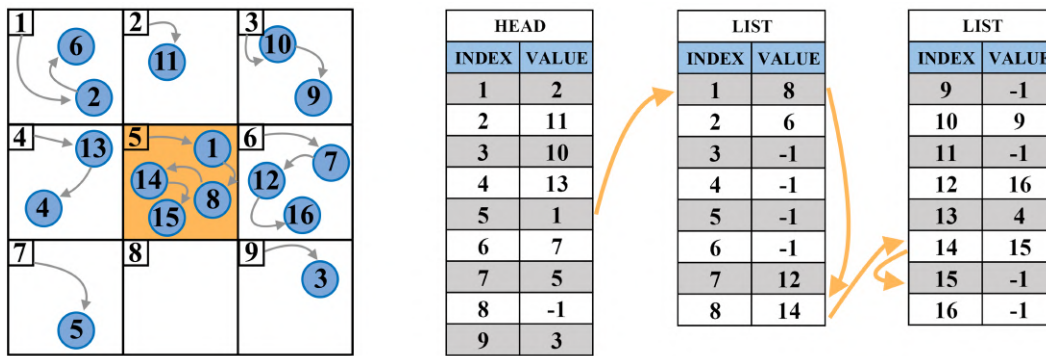


Figura A.2. Schema nel caso bidimensionale del funzionamento delle linked cell list. Le frecce descrivono il processo per la cella centrale.

Nel programma, la funzione **MAPS**, chiamata una volta all'inizio della simulazione, traccia la mappa delle 26 celle vicine ad ogni cella. Subito dopo la funzione **LINKS** individua la cella di appartenenza per ogni particella, andando a compilare i due array **LIST** e **HEAD**. Successivamente se una particella nello spostamento uscisse dalla sua cella iniziale si utilizza la funzione **list - remove** per rimuovere la particella dalla cella iniziale e si utilizza **list - add** per aggiungere la particella nella sua nuova cella.

Appendice B

Rilevamento di collisioni - 3D

Minkowski Portal Refinement

Si discute il *Minkowski Portal Refinement* nel caso tridimensionale. Si inizia definendo lo pseudo-codice dell'algoritmo *XenoCollide*, mostrato nel listato B.1.

```
// Fase 1: Ricerca dei portali
find_origin_ray();
find_candidate_portal();
while (origin ray does not intersect candidate) {
    choose_new_candidate();
}

// Fase 2: Perfezionamento del portale
while (true) {
    if (origin inside portal) return hit;
    find_support_in_direction_of_portal();
    if (origin outside support plane) return miss;
    if (support plane close to portal) return miss;
    choose_new_portal();
}
```

Listing B.1. Pseudo-codice dell'algoritmo *XenoCollide*.

Si procede quindi a descrivere le varie funzioni utilizzate nel programma per spiegare anche la metodologia con cui l'algoritmo trova le collisioni tra poliedri.

Funzione *find_origin_ray()*

L'algoritmo inizia cercando un punto all'interno della differenza di Minkowski $B - A$, in generale può essere presa la differenza dei centri di massa dei due oggetti V_0 (figura B.1a). Se il vettore origine, che collega V_0 all'origine, passa attraverso una superficie di $B - A$ prima di arrivare all'origine, allora quest'ultima sarà al di fuori della differenza di Minkowski. Altrimenti se incontra prima l'origine, questa sarà contenuta, e quindi i due oggetti avranno un'intersezione.

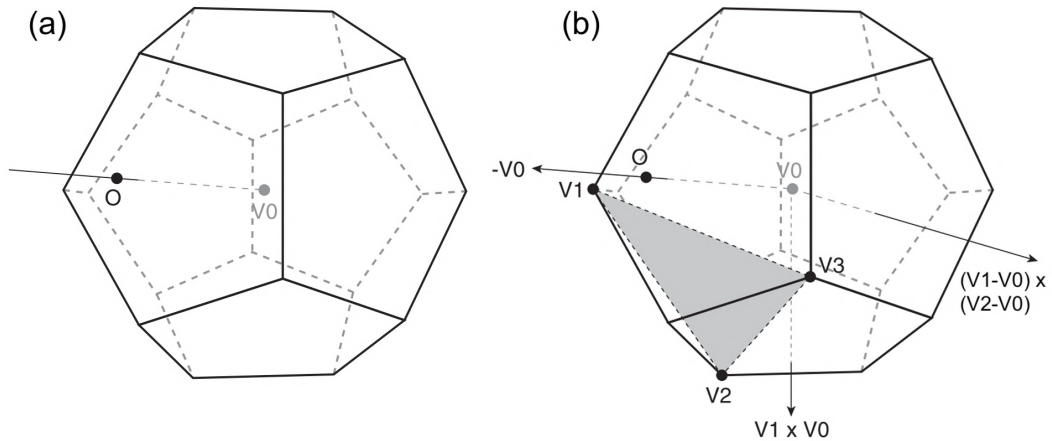


Figura B.1. (a) Rappresentazione del vettore origine su una differenza di Minkowski che collega V_0 all'origine. (b) Ricerca del portale triangolare nel poliedro. Immagina tratta da [22].

Funzione *find_candidate_portal()*

In questa funzione si utilizza la mappatura di supporto di $B - A$ per trovare tre punti non allineati che formano un portale triangolare in cui può o meno passare il vettore origine. Come mostrato in figura B.1b.

Ciclo *while (origin ray does not intersect candidate)*

Si controlla se il vettore origine intersechi il portale triangolare, per farlo si controlla se l'origine è contenuta all'interno dei tre piani formati dai lati triangolari (V_0, V_1, V_2) , (V_0, V_2, V_3) e (V_0, V_3, V_1) . Se è contenuta questo portale è un buon candidato altrimenti bisogna riprovare con un altro portale.

Funzione *choose_new_candidate()*

Se l'origine è al di fuori dei tre piani formati si usa la normale al piano in cui l'origine è fuori per trovare il terzo punto del nuovo portale, il processo è mostrato in figura B.2a. Il processo si ripeterà finché l'origine non sarà contenuta nel portale.

Check *if(origin inside portal) return hit*

I punti V_0, V_1, V_2 e V_3 formano un tetraedro, che, per costruzione, sarà totalmente contenuto in $B - A$. Di conseguenza se il tetraedro contiene l'origine allora anche la differenza di Minkowski conterrà l'origine, in questo caso il programma si interromperà e riporterà una intersezione tra A e B .

Funzione *find_support_in_direction_of_portal()*

A questo punto ci sono due casi possibili, l'origine può essere ancora contenuta in $B - A$ ma è al di fuori del tetraedro costruito o l'origine è al di fuori della differenza

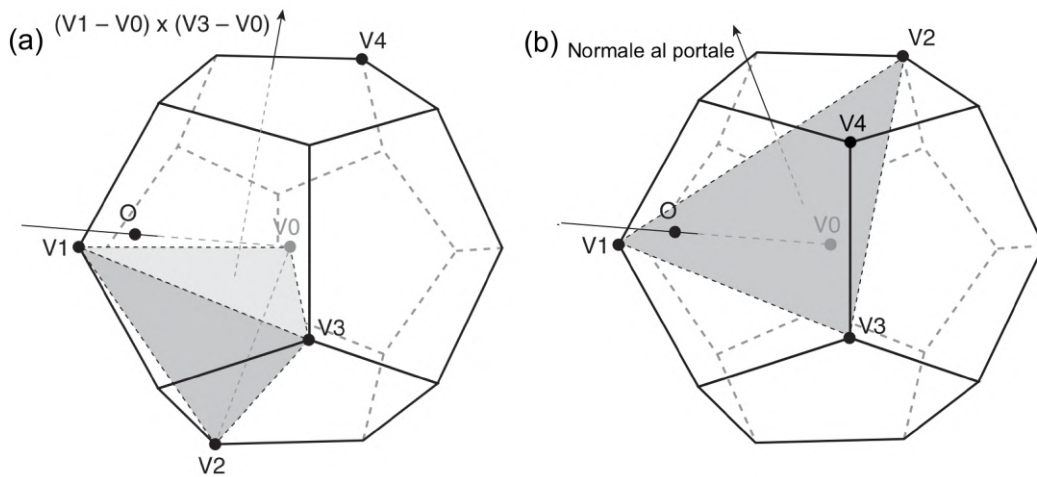


Figura B.2. (textbfa) Ricerca del nuovo portale triangolare, trovata la direzione si utilizza la mappatura del supporto per trovare il vertice V_2 . (textbfb) Il nuovo portale formato con la relativa normale. Immagina tratta da [22].

di Minkowski. Per distinguere i due casi si cerca il nuovo punto di supporto costruito sulla normale che guarda al di fuori di $B - A$. Come mostrato in figura B.2 (b).

Check if (origin outside support plane) return miss

Se l'origine è al di fuori del nuovo piano costruito sulla normale con il nuovo punto di supporto allora l'origine sarà al di fuori di $B - A$ e l'algoritmo riporta i due poliedri come separati. In questo caso il programma cercherà un nuovo portale siccome la non sovrapposizione in un portale non assicura che l'origine non sia contenuta in un altro portale.

Funzione choose_new_portal()

Bisogna rifinire la ricerca dell'origine, quindi trovare un portale che sia più vicino alla superficie di $B - A$. Si consideri ora la normale al portale ($V1, V2, V3$), mostrata in figura B.2 (b), e si utilizzi la funzione mappatura di supporto su questa direzione trovando il punto $V4$. Si consideri quindi il tetraedro formato dai punti $V1, V2, V3$ e $V4$, mostrato in figura B.3 (a), il vettore origine passa attraverso il portale ($V1, V2, V3$), quindi è garantito che il vettore origine uscirà attraverso una delle altre tre facce del tetraedro, bisogna individuare quale. Per farlo si confronta la posizione dell'origine con i tre piani formati dai vertici del tetraedro: ($V4, V0, V1$), ($V4, V0, V2$) e ($V4, V0, V3$). Si seleziona quindi l'unico piano in cui l'origine e il centro del tetraedro costruito non si trovino dallo stesso lato rispetto al piano considerato. Il nuovo portale sarà definito dalla faccia del tetraedro che forma il piano selezionato, come mostrato in figura B.3 (b) ($V1, V3, V4(old)$).

Check if (support plane close to portal) return miss

Aumentando il numero delle iterazioni, il portale si avvicina sempre di più alla superficie $B - A$. C'è da aprire una parentesi, perché, se l'oggetto di cui si cerca di

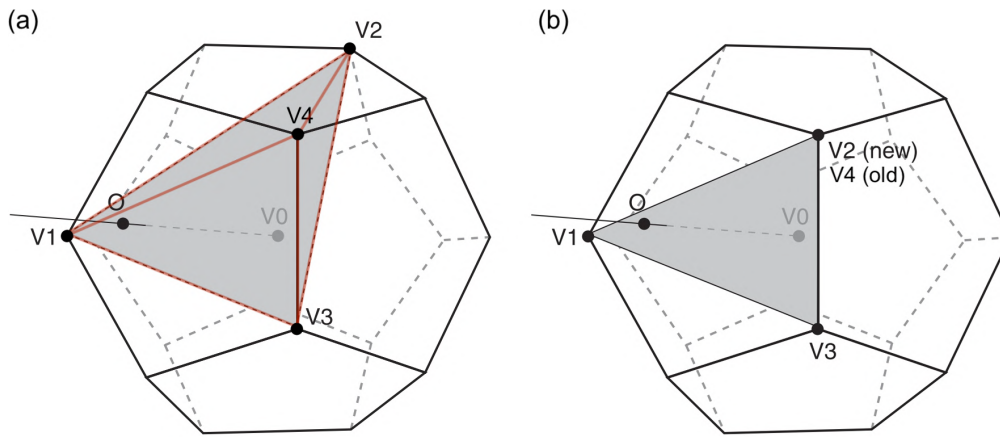


Figura B.3. (a) Nel caso in cui l'origine non sia inclusa nel tetraedro si utilizza la funzione di supporto sulla normale del portale uscente del poliedro, trovando così il punto V_4 . Nell'immagine viene evidenziato il tetraedro formato dai punti V_1 , V_2 , V_3 e V_4 . (b) Si costruisce il successivo portale utilizzando un nuovo vertice V_2 . Immagina tratta da [22].

trovare le intersezioni ha una faccia curva, l'origine potrebbe essere infinitesimamente vicina a essa. Per questo è necessario definire una tolleranza sulla distanza tra portale e superficie tonda, superata questa l'algoritmo termina. In questo caso, in ambito fisico si preferisce riportare un'intersezione di due oggetti così da evitare sovrapposizioni.

Output

L'algoritmo continuerà a girare finché una di queste condizioni non è soddisfatta:

- L'origine è contenuta in un portale, il programma riporta una intersezione.
- L'origine è all'esterno del piano di supporto, il programma riporta una non intersezione.
- La distanza tra il portale e il suo piano di supporto parallelo è inferiore a una tolleranza, il programma restituisce una intersezione o una non intersezione a seconda dell'applicazione.

Bibliografia

- [1] D.J. Adams. Grand canonical ensemble monte carlo for a lennard-jones fluid. *Molecular Physics*, 29(1):307–311, 1975.
- [2] Tommaso Bellini, Roberto Cerbino, and Giuliano Zanchetta. Dna-based soft phases. *Topics in current chemistry*, 318:225–79, 08 2011.
- [3] Anne M. Bohle, Robert Holyst, and Thomas Vilgis. Polydispersity and ordered phases in solutions of rodlike macromolecules. *Phys. Rev. Lett.*, 76:1396–1399, 1996.
- [4] Peter Bolhuis and Daan Frenkel. Tracing the phase boundaries of hard spherocylinders. *The Journal of Chemical Physics*, 106(2):666–687, 1997.
- [5] Y. Bouligand. Twisted fibrous arrangements in biological materials and cholesteric mesophases. *Tissue and Cell*, 4(2):189–217, 1972.
- [6] Marvin H. Caruthers. Gene synthesis machines: Dna chemistry and its uses. *Science*, 230(4723):281–285, 1985.
- [7] S. Chandrasekhar. *Liquid Crystals*. Cambridge University Press, 1992.
- [8] Nanette J. Pazdernik David P. Clark and Michelle R. McGehee. *Molecular Biology*. Elsevier Academic Press, 2019.
- [9] P. J. de Gennes. *The Physics of Liquid Crystal*. Oxford University Press, 1974.
- [10] Cristiano De Michele, Tommaso Bellini, and Francesco Sciortino. Self-assembly of bifunctional patchy particles with anisotropic shape into polymers chains: Theory, simulations, and experiments. *Macromolecules*, 45(2):1090–1106, 2012.
- [11] Cristiano De Michele, Lorenzo Rovigatti, Tommaso Bellini, and Francesco Sciortino. Self-assembly of short dna duplexes: from a coarse-grained model to experiments through a theoretical link. *Soft Matter*, 8:8388–8398, 2012.
- [12] Ingo Dierking. *Textures of Liquid Crystals*. WILEY-VCH Verlag GmbH & Co. KGaA, 2003.
- [13] David H. Eberly. Testing for intersection of convex objects: The method of separating axes. 2001.
- [14] D.H. Eberly. *3D Game Engine Design: A Practical Approach to Real-Time Computer Graphics*. Morgan Kaufmann Publishers, 2001.

- [15] G. Pesole F. Amaldi, P. Benedetti and P. Plevani. *Biologia Molecolare*. Casa Editrice Ambrosiana, 2018.
- [16] Daniel Fiser. Xenocollide mpr algorithm. <https://github.com/danfis/libccd>, 2012. Copyright (c)2010-2012 Daniel Fiser <danfis@danfis.cz>, Intelligent and Mobile Robotics Group, Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague. All rights reserved.
- [17] G. Friedel. Les états mésomorphes de la matière. *Ann. Phys.*, 9(18):273–474, 1922.
- [18] Manning Gerald S. The persistence length of dna is reached from the persistence length of its null isomer through an internal electrostatic stretching force. *Biophys J*, 91(10):3607–3616, 2006.
- [19] E.G. Gilbert, D.W. Johnson, and S.S. Keerthi. A fast procedure for computing the distance between complex objects in three-dimensional space. *IEEE Journal on Robotics and Automation*, 4(2):193–203, 1988.
- [20] James J. DeLuca Gregory A. DiLisi. *An Introduction to Liquid Crystals*. Morgan & Claypool, 2019.
- [21] Alexander Yu. Grosberg and Alexei R. Khokhlov. *Statistical Physics of Macromolecules*. Grosberg, 1994.
- [22] S. Jacobs. *Game Programming Gems 7*. Course Technology, Cengage Learning, 2008.
- [23] Tony Z. Jia, Tommaso Bellini, Noel Clark, and Tommaso P. Fraccia. A liquid crystal world for the origins of life. *Emerging Topics in Life Sciences*, 6(6):557–569, 11 2022.
- [24] Donald E. Knuth. *The Art of Computer Programming*. Addison Wesley, 1997.
- [25] Tatiana Kuriabova, M. D. Betterton, and Matthew A. Glaser. Linear aggregation and liquid-crystalline order: comparison of monte carlo simulation and analytic theory. *J. Mater. Chem.*, 20:10366–10383, 2010.
- [26] Sven T. Lagerwall. Comments on liquid crystal terminology, nomenclature and conventions. *Ferroelectrics*, 85(1):497–501, 1988.
- [27] Otto Lehmann. Über fließende krystalle. *Zeitschrift für Physikalische Chemie*, 4U:462 – 472, 1889.
- [28] Jinsen Li, Jared M. Sagendorf, Tsu-Pei Chiu, Marco Pasi, Alberto Pérez, and Remo Rohs. Expanding the repertoire of dna shape features for genome-scale studies of transcription factor binding. *Nucleic Acids Research*, 45:12877 – 12887, 2017.
- [29] Li Li, Suzette Pabit, Jessica Lamb, Hye Yoon Park, and Lois Pollack. Closing the lid on dna end-to-end stacking interactions. *Applied physics letters*, 92:223901–2239013, 07 2008.

- [30] Francoise Livolant and Amélie Leforestier. Condensed phases of dna: Structures and phase transitions. *Progress in Polymer Science*, 21(6):1115–1164, 1996.
- [31] Yongjun Lu, Brock Weers, and Nancy C. Stellwagen. Dna persistence length revisited. *Biopolymers*, 61(4):261–275, 2002.
- [32] K. Merchant and R.L. Rill. Dna length and concentration dependencies of anisotropic phase transitions of dna solutions. *Biophysical Journal*, 73(6):3154–3163, 1997.
- [33] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.
- [34] Nicholas Metropolis and S. Ulam. The monte carlo method. *Journal of the American Statistical Association*, 44(247):335–341, 1949. PMID: 18139350.
- [35] Michi Nakata, Giuliano Zanchetta, Brandon D. Chapman, Christopher D. Jones, Julie O. Cross, Ronald Pindak, Tommaso Bellini, and Noel A. Clark. End-to-end stacking and liquid crystal condensation of 6 to 20 base pair dna duplexes. *Science*, 318(5854):1276–1279, 2007.
- [36] Khanh Thuy Nguyen, Francesco Sciortino, and Cristiano De Michele. Self-assembly-driven nematization. *Langmuir*, 30(16):4814–4819, 2014. PMID: 24701976.
- [37] Shuichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2):255–268, 1984.
- [38] Paul A. O’Brien, Michael P. Allen, David L. Cheung, Matthew Dennison, and Andrew Masters. Elastic constants of hard thin platelets by monte carlo simulation and virial expansion. *Phys. Rev. E*, 78:051705, 2008.
- [39] Lars Onsager. The effects of shape on the interaction of colloidal particles. *Annals of the New York Academy of Sciences*, 51(4):627–659, 1949.
- [40] Melissa E. O’Neill. Pcg: A family of simple fast space-efficient statistically good algorithms for random number generation. 2014.
- [41] D.H. Eberly P. Schneider. *Geometric Tools for Computer Graphics*. Morgan Kaufmann Publishers, 2002.
- [42] Linus Pauling and Robert B. Corey. A proposed structure for the nucleic acids. *Proc. Natl. Acad. Sci.*, 39(2):84–97, 1953.
- [43] Jure Piškur and Allan Rupprecht. Aggregated dna in ethanol solution. *FEBS Letters*, 375(3):174–178, 1995.
- [44] Matthew R. Redinbo, Lance Stewart, Peter Kuhn, James J. Champoux, and Wim G. J. Hol. Crystal structures of human topoisomerase i in covalent and noncovalent complexes with dna. *Science*, 279(5356):1504–1513, 1998.

- [45] Friedrich Reinitzer. Beiträge zur kenntniss des cholesterins. *Monatshefte für Chemie und verwandte Teile anderer Wissenschaften*, 9:421–441, 1888.
- [46] Friedrich Reinitzer. Contributions to the knowledge of cholesterol. *Liquid Crystals*, 5(1):7–18, 1989.
- [47] Conmar Robinson. Liquid-crystalline structures in polypeptide solutions. *Tetrahedron*, 13(1):219–234, 1961.
- [48] H. Sackmann and D. Demus. The polymorphism of liquid crystals. *Molecular Crystals*, 2(1-2):81–102, 1966.
- [49] John Santalucia. A unified view of polymer, dumbbell, and oligonucleotide dna nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 95:1460–5, 03 1998.
- [50] John SantaLucia and Donald Hicks. The thermodynamics of dna structural motifs. *Annual Review of Biophysics and Biomolecular Structure*, 33(1):415–440, 2004. PMID: 15139820.
- [51] Francesco Sciortino, Emanuela Bianchi, Jack F. Douglas, and Piero Tartaglia. Self-assembly of patchy particles into polymer chains: A parameter-free comparison between wertheim theory and monte carlo simulation. *The Journal of Chemical Physics*, 126(19):194903, 2007.
- [52] Jonathan V. Selinger and Robijn F. Bruinsma. Hexagonal and nematic phases of chains. i. correlation functions. *Phys. Rev. A*, 43:2910–2921, 1991.
- [53] Gregory P. Smith, Tommaso P. Fraccia, Marco Todisco, Giuliano Zanchetta, Chenhui Zhu, Emily Hayden, Tommaso Bellini, and Noel A. Clark. Backbone-free duplex-stacked monomer nucleic acids exhibiting watson–crick selectivity. *Proceedings of the National Academy of Sciences*, 115(33):E7658–E7664, 2018.
- [54] Gary Snethen. Xenocollide website. <http://www.xenocollide.com>, 2007.
- [55] Bernard Tinland, Alain Pluen, Jean Sturm, and Gilbert Weill. Persistence length of single-stranded dna. *Macromolecules*, 30(19):5763–5765, September 1997.
- [56] Fabio Tombolato and Alberta Ferrarini. From the double-stranded helix to the chiral nematic phase of b-dna: A molecular model. *The Journal of chemical physics*, 122:54908, 03 2005.
- [57] Mark Tuckerman. Statistical mechanics: Theory and molecular simulation. 01 2001.
- [58] J.D. Watson and F.H.C. Crick. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, 1953.
- [59] Richard A. Wing, Horace R. Drew, Tsunehiro Takano, Chris Allen Broka, Shoji Tanaka, Keiichi Itakura, and Richard E. Dickerson. Crystal structure analysis of a complete turn of b-dna. *Nature*, 287:755–758, 1980.

- [60] Peter Yakovchuk, Ekaterina Protozanova, and Maxim D. Frank-Kamenetskii. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Research*, 34(2):564–574, 01 2006.
- [61] Giuliano Zanchetta. Spontaneous self-assembly of nucleic acids: Liquid crystal condensation of complementary sequences in mixtures of dna and rna oligomers. *Liquid Crystals Today*, 18, 09 2009.
- [62] Giuliano Zanchetta and Roberto Cerbino. Exploring soft matter with x-rays: from the discovery of the DNA structure to the challenges of free electron lasers. *Journal of Physics: Condensed Matter*, 22(32):323102, 2010.