



SAPIENZA  
UNIVERSITÀ DI ROMA

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

**Aggregazione di Doppi Filamenti di DNA:  
Ruolo della Conformazione  
della Struttura a Doppia Elica**

---

**Relatore:** Prof. CRISTIANO DE MICHELE

**Laureando:** DANIELE ANCORA  
**Matricola:** 1096696



*“Amusons-nous sur la terre et sur l’onde*

*Malheureux, qui se fait un nom!*

*Richesses, honneurs, faux éclat de ce monde,*

*Tout n’est que boules de savon”*

Per Te, La Soffleuse de Savon - François Boucher.



Ai Miei Genitori, a Mia Sorella, a Mia Nonna che c'è

ed agli altri Nonni, che porto sempre con me.

Ed alla Mia Famiglia, da sempre acqua nel deserto.



# Indice

<b>I</b>	<b>Background Scientifico</b>	<b>18</b>
<b>1</b>	<b>Il ruolo del DNA nella Fisica della Materia Soffice</b>	<b>19</b>
1.1	Primi passi nell'investigazione della struttura dell'acido nucleico . . . . .	19
1.2	Il processo di "Self-Assembly" . . . . .	22
<b>2</b>	<b>Struttura del DNA</b>	<b>26</b>
2.1	Composizione Chimica . . . . .	26
2.1.1	Basi Azotate . . . . .	28
2.1.2	Zucchero Pentoso . . . . .	29
2.1.3	Gruppo Fosfato . . . . .	30
2.2	L'accoppiamento delle basi azotate . . . . .	31
2.3	Morfologia Strutturale della Doppia Elica . . . . .	34
2.4	Nomenclatura: il formato *.pdb . . . . .	36
<b>3</b>	<b>Simulazione e Modellizzazione Computazionale del DNA</b>	<b>38</b>
3.1	Approccio "Bottom-Up" . . . . .	39
3.2	Approccio "Top-Down" . . . . .	40
3.3	Il nostro approccio . . . . .	41
<b>II</b>	<b>Dinamica Molecolare full-atom</b>	<b>45</b>
<b>4</b>	<b>Cenni di teoria sulla MD-Simulation</b>	<b>46</b>
4.1	Software GROMACS . . . . .	48
4.2	T trattazione delle forze e delle interazioni tra particelle . . . . .	51
4.2.1	ffAMBER03 . . . . .	52
4.2.2	Modello TIP3P . . . . .	53
4.2.3	T trattazione del "cut-off" delle interazioni . . . . .	55
4.3	Set-up dei parametri della simulazione . . . . .	56
4.4	Preparazione del sistema . . . . .	58

4.5	Minimizzazione dell'energia potenziale . . . . .	60
<b>5</b>	<b>Parallelizzazione su SuperCalcolatori</b>	<b>63</b>
5.1	Parallel-Computing con MPI . . . . .	63
5.2	Parallelizzazione con Cluster Matrix . . . . .	65
5.3	Parallelizzazione con JUROPA/HPC-FF . . . . .	67
5.4	Parallelizzazione con Fermi BlueGene/Q . . . . .	69
<b>6</b>	<b>Parallelizzazione su GPU</b>	<b>72</b>
6.1	Parallel-Computing con CUDA . . . . .	72
6.2	Parallelizzazione su nVidia Tesla . . . . .	74
<b>7</b>	<b>Analisi dati</b>	<b>76</b>
7.1	Simulazioni di DNA in soluzione acquosa . . . . .	76
7.2	Equilibratura della struttura terziaria dell'acido nucleico . . . . .	80
7.3	Convinciamoci dell'equilibratura . . . . .	83
<b>8</b>	<b>Definizioni delle osservabili</b>	<b>85</b>
8.1	Il "bending" assiale . . . . .	87
8.2	Baricentri dei gruppi fosforici . . . . .	90
<b>9</b>	<b>Analisi delle simulazioni single-DNA</b>	<b>95</b>
9.1	Effetto di taglia del sistema . . . . .	95
<b>10</b>	<b>Analisi delle simulazioni multi-DNA</b>	<b>99</b>
10.1	Angolazioni e distanza tra Basi . . . . .	99
10.2	Bending assiale . . . . .	104
10.3	"Bending" fosforico . . . . .	108
<b>III</b>	<b>Simulazione Monte Carlo di cilindri piegati</b>	<b>112</b>
<b>11</b>	<b>Teoria sulla dinamica Monte Carlo</b>	<b>113</b>
11.1	Simulazioni Monte Carlo NVT . . . . .	115
11.2	Simulazioni Monte Carlo NPT . . . . .	118
11.3	Teoria sulla termodinamica di Cilindri Rigidi con siti attrattivi . . . . .	120
11.3.1	Fase isotropica . . . . .	123
11.3.2	Fase nematica . . . . .	124
11.3.3	Regione di coesistenza . . . . .	125

<b>12 Simulazioni MC di Cilindri Piegati</b>	<b>126</b>
12.1 Modellizzazione di DNA piegati . . . . .	126
12.2 Volume degli oggetti . . . . .	128
12.3 Confronto con i dati sperimentali . . . . .	128
12.4 Definizione delle osservabili di nematizzazione . . . . .	130
12.5 Set-up sperimentale del sistema . . . . .	131
<b>13 Analisi dati</b>	<b>135</b>
13.1 Equilibratura del sistema . . . . .	136
13.2 Il primo gruppo . . . . .	139
13.3 Il secondo gruppo . . . . .	142
13.4 Dipendenza della concentrazione critica dalla conformazione della molecola . . .	144



# Introduzione

Il settore derivato dalla Fisica della Materia Condensata che si occupa dello studio delle proprietà collettive di aggregati macroscopici formati da “building blocks” di dimensioni mesoscopiche, in termini delle interazioni che intercorrono tra i loro stessi costituenti, viene contraddistinto con il nome di *Fisica della Materia Soffice*. Nel 1991 Pierre Guilles de Gennes, fisico francese impegnato anche nello studio dei materiali semiconduttori, venne insignito del Premio Nobel per la fisica per aver scoperto che lo studio dei fenomeni legati al comportamento del parametro d’ordine, in sistemi termodinamici semplici, poteva essere esteso e generalizzato a sistemi formati da oggetti strutturalmente più complessi, come ad esempio i polimeri ed i cristalli liquidi.

Il lavoro di de Gennes contraddistingue la nascita di quella che viene definita “Soft Matter” [1] e segna un punto di svolta nello studio e nel controllo di interazioni tra grandi aggregati molecolari facilmente deformabili da fluttuazioni o stress termici, fenomeno di grande interesse a causa della crescente necessità di sviluppare materiali innovativi.

Sostanze liquide, colloidali, ma anche polimeri, schiume, gel, materiali granulari e biologici sono gli oggetti coinvolti in questo studio [2, 3] che cerca di spiegare la ragione della presenza di elementi comuni nei comportamenti fisici collettivi di tali materiali, osservabili a scale di energia grossomodo comparabili con l’energia termica a temperatura ambiente ( $k_B T$ ).

Le proprietà ottiche e meccaniche dei materiali “soffici” sono tipicamente determinate dalla loro organizzazione in aggregati costituiti da “building blocks”, i blocchi fondamentali che formeranno l’aggregato, caratterizzati da dimensioni che variano da qualche nanometro ad alcuni micron. Il comportamento macroscopico collettivo di questi materiali può esibire una grande varietà di livelli di ordinamento: partendo da strutture a basso grado di organizzazione, in cui



Figura 1: Pierre Guilles de Gennes, padre fondatore della Fisica della Materia Soffice.

l'ordine traslazionale è esibito solamente a livello locale, fino ad arrivare a strutture ad alto grado di ordinamento, ad esempio mediante la formazione di strutture cristalline regolari.

Il processo che sta alla base di tale fenomeno viene denominato "*Self-Assembly*", traducibile in italiano con i termini "auto-assemblamento" oppure "auto-organizzazione", ovvero la capacità di tali composti di formare strutture ordinate minimizzando la propria energia libera come conseguenza di interazioni tra gli stessi costituenti. La caratteristica peculiare di ciò risiede nel fatto che questa capacità di aggregazione spontanea può essere controllata esternamente mediante variabili macroscopiche globali, quali ad esempio temperatura, pressione o concentrazione oppure direttamente mediante opportune scelte nella composizione chimica dei costituenti e delle interazioni tra gli stessi. Intervenendo su questi semplici parametri, quindi, riusciamo ad arrivare ad un elevato grado di ordinamento strutturale microscopico e la vastità delle interazioni sinora studiate coinvolte nei processi di auto-aggregazione, fornisce ai ricercatori uno strumento sempre più potente per lo sviluppo di nuovi tipi di materiali, stimolando al tempo stesso la ricerca e lo studio dei meccanismi alla base di tali fenomeni.

In questo contesto scientifico, occupano un ruolo di grande rilievo i processi legati alle interazioni biologiche per via delle peculiari caratteristiche chimico-fisiche coinvolte: tali processi sono infatti caratterizzati da specificità, direzionalità, reversibilità e da una vasta gamma di possibilità di manipolazione delle strutture organiche ad esse connesse, caratteristiche che ben si prestano a molteplici applicazioni nella fisica della materia soffice.

In particolare l'interesse scientifico per la molecola organica per eccellenza, ovvero l'*acido desossiribonucleico* conosciuto con il noto acronimo di DNA, ha spinto i ricercatori di tutto il mondo a sondare con grande attenzione e minuziosità un numero sempre più vasto di interazioni ad esso legate. L'interesse dietro questo particolare biopolimero non è dato solamente dal fatto che con tale molecola viene salvato e tramandato il codice genetico di qualsiasi specie vivente, ma anche dalla sua particolare conformazione a doppia elica e dalle possibilità di variarne lunghezza e tipologia di sequenza. Attraverso particolari processi chimici e di bio-ingegnerizzazione, il DNA può essere letteralmente costruito da zero decidendone a priori tutte le sue caratteristiche chimico-strutturali, le quali giocano un ruolo cruciale nella caratterizzazione della termodinamica, nelle proprietà meccaniche e nella geometria del sistema, rendendo quindi il DNA un candidato perfetto per lo studio delle interazioni tipiche della "soft-matter". Questo processo di ingegnerizzazione degli acidi nucleici ha aperto nuove possibilità di studio, in particolare dei fenomeni connessi alla morfologia strutturale dell'acido nucleico in funzione della sequenza di cui è costituito, permettendo l'analisi di sistemi puri costituiti da doppi filamenti aventi determinate sequenze di basi azotate.

Il "tuning" fine delle interazioni durante il processo di aggregazione dei polimeri ha permesso di arrivare ad un elevato grado di ingegnerizzazione delle strutture create, consentendo ai ricercatori di intervenire sulla geometria di aggregazione e permettendo un controllo pressoché totale del processo. Nei lavori di Luo et al. [7, 8] la scelta di una particolare geometria

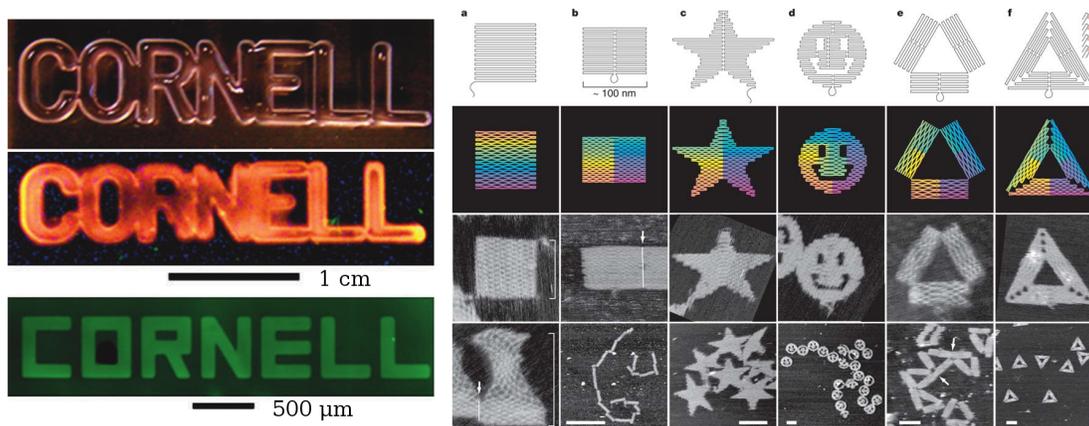


Figura 2: Le scritte “CORNELL” in “hydrogel” formato da DNA, ad opera del gruppo di Luo [8], illuminate mediante coloranti fluorescenti. Le dimensioni macroscopiche degli oggetti ( $1\text{cm} \div 500\mu\text{m}$ ) ci mostrano quanto sia potente il processo di “self assembly” nel formare aggregati ordinati su grande scala. A sinistra i particolari origami larghi circa  $\sim 100\text{nm}$  ideati da Rothemund [9] attraverso il processo di “folding” del singolo filamento di acido nucleico.

nella costruzione di dendrimeri formati da filamenti di DNA ha permesso a tali molecole l’aggregazione catalizzata da opportuni enzimi in aggregati supra-molecolari (hydrogel), attraverso i quali è stato possibile realizzare delle suggestive scritte fluorescenti. Ancora più particolare lo studio di Rothemund [9] nel quale, sfruttando il processo di “folding” dei singoli filamenti di acido nucleico ed attraverso il posizionamento sequenziale di determinate nucleobasi, il controllo del processo di “base pairing” ha consentito la creazione di origami dalle forme più disparate.

Inoltre il processo di “self-assembly” di biopolimeri di DNA ha attirato l’attenzione dei teorici e degli sperimentali per il suo potenziale nella fisica della materia soffice dal punto di vista termodinamico, come recentemente mostrato da Bellini et al. [10]. In questo lavoro viene messa in particolare risalto la formazioni di fasi liquido-cristalline in sistemi costituiti da colloidi rivestiti da doppie eliche di DNA. Tali particelle possono essere ottenute mediante il processo di funzionalizzazione di biopolimeri su normali particelle colloidali ideato da Mirkin [11] e successivamente studiato da Glotzer [12] nel caso dell’oro, oppure anche utilizzando altri materiali, quali argento o polistirene. Inoltre il processo di auto-organizzazione può essere usato a proprio vantaggio per ottenere una serie di costrutti con varie proprietà particolari, ad esempio decidendone a priori la loro geometria o forma come mostrato nel pionieristico lavoro di Seeman [13], il primo a sfruttare tali incredibili caratteristiche fisiche.

Essendo il DNA assimilabile sostanzialmente ad un cilindro rettilineo ed avendo una struttura abbastanza rigida con una lunghezza di persistenza pari a circa  $125\text{bp}$ , risultò naturale scoprire che tale acido nucleico potesse transire in una fase nematica ad ordinamento colonnare

a determinate concentrazioni critiche, comportandosi come un cristallo liquido.

Un grande contributo allo studio della fase liquido-cristallina di sistemi termodinamici composti da segmenti di DNA viene fornito dal modello proposto da Onsager [14] nel 1949, che mostrò teoricamente che il solo contributo di interazione del volume escluso di particelle allungate, assimilabili a cilindri rigidi (interazioni “hard-core”), era sufficiente a produrre un allineamento macroscopico delle particelle in soluzione. Successivamente Bolhuis-Frenkel [15] caratterizzarono completamente il diagramma di fase attraverso lo studio computazionale del sistema “rod-like” proposto da Onsager in cui le particelle in soluzione sono viste come delle barre allungate di forma cilindrica. In questo modo riuscirono a prevedere l’esistenza di un limite inferiore all’“aspect ratio”, ovvero il rapporto tra lunghezza e diametro del cilindro che lo approssima  $X_0 = L/D$ , che permette la transizione di fase: a densità inferiori al valore critico  $X_0^{critical} \simeq 4.7$ , cioè corrispondente circa ad un DNA lungo  $N = 28$  basi azotate, il contributo del solo volume escluso non è sufficiente a permettere la transizione nematica al sistema.

L’accordo sperimentale con il modello di Onsager-Bolhuis-Frenkel [14, 15] faceva ipotizzare che il fenomeno della nematizzazione di polimeri “rod-like” potesse essere spiegato solamente in termini del volume escluso, ma recentemente il lavoro di Nakata e Zanchetta [16] ha portato alla luce tale transizione di fase anche in segmenti molto più corti da quanto previsto dalla teoria: essi infatti hanno osservato il fenomeno della nematizzazione per doppie eliche di DNA corte le cui sequenze palindrome erano costituite da 6 a 20 basi azotate. A complicare ulteriormente il quadro della situazione, Nakata et al. nel loro successivo lavoro [17], hanno messo in evidenza altre nuove proprietà peculiari di queste fasi: la concentrazione a cui i vari sistemi transiscono dipende criticamente dalla sequenza con cui viene preparata la soluzione, come pure la chiralità globale per la fase nematica ottenuta. Non potendo il volume escluso essere il principale responsabile di questo fenomeno, tale scoperta ha spalancato le porte nello studio di particolari interazioni già note in altri sistemi biologici, le forze di “stacking”, forze di natura attrattiva che si instaurerebbero tra le basi terminali dei segmenti

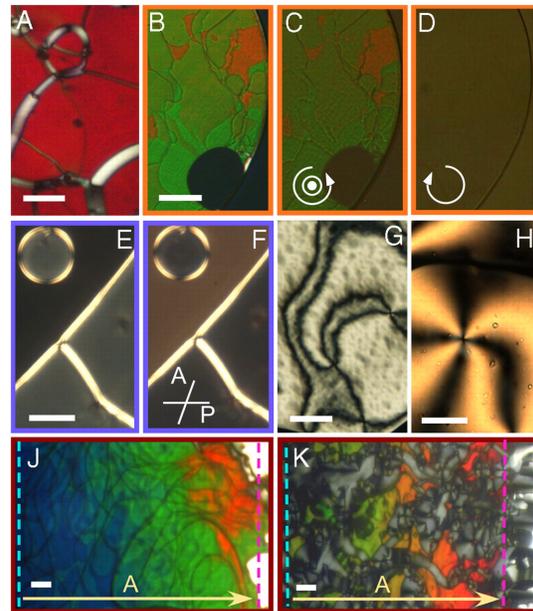


Figura 3: Immagine delle fasi nematiche di sequenze corte di DNA riscontrate nel lavoro di Nakata e Zanchetta [16] (PTOM microscopic textures).

di DNA e possibili responsabili dell'incollamento e della successiva formazione della fase liquido-cristallina. Come stimato sperimentalmente dai lavori di Bellini et al. [17] le energie coinvolte in tale fenomeno risultano essere dell'ordine di  $\Delta E_S \simeq 4 \div 8 k_B T$ .

In questo contesto, il lavoro teorico e computazionale di De Michele et al. [18] riesce a dare una spiegazione al fenomeno di nematizzazione riscontrato da Nakata in tali sequenze corte, approssimando il DNA a degli ellissoidi rigidi di rotazione ("Hard Ellipsoid", HE) di forma cilindrica caratterizzati dall'"aspect ratio"  $X_0$  e da due siti attrattivi sulle estremità che modellizzano le forze di "stacking". Grazie a questo modello "coarse-grained" è stato possibile mostrare, ricostruendo i diagrammi di fase, che tali sistemi riescono a nematizzare anche con "aspect ratio" inferiori a  $X_0 < 3.0$  contrariamente a quanto previsto dal modello OBF. Inoltre De Michele et al. in un successivo lavoro [19] riescono a fornire una buona stima delle forze che intervengono tra coppie di DNA usando delle superquadriche (SQ) al posto dei cilindri, ottenendo valori dell'energia di "stacking" che si attestano intorno a  $\beta \Delta E_s \simeq 5.56 \div 8.33$  compatibilmente con quanto ottenuto dai dati sperimentali [16].

In un successivo lavoro Rovigatti e De Michele [20] provano ancora a studiare, implementando il modello di Ouldridge mediante dinamica molecolare, la termodinamica di sistemi di DNA corti approssimando ogni singolo nucleotide con strutture "coarse grained" più complesse. Alla luce di quanto visto in [19], l'energia libera  $G_{bind} = \Delta E_s - T \Delta S_s$  responsabile del fenomeno risulta essere attorno  $G_{bind} = -(0.4 \div 2.4) \text{ kcal/mol}$ .

Nell'ottica di studiare sempre più nel dettaglio il fenomeno di aggregazione di filamenti di DNA corti, il gruppo di Maffeo [21] ha proposto invece un approccio "full-atom" all'analisi del problema, effettuando un calcolo dell'energia di "stacking" mediante l'uso della dinamica molecolare applicata allo studio di un sistema costituito da due segmenti di DNA. In soluzione acquosa hanno osservato l'adesione spontanea delle doppie eliche mediante forze di stacking tra le basi estremali, mettendo alla luce il processo di "self-assembly" nella singola coppia, ottenendo per le forze responsabili del fenomeno energie libere stimate di  $G_{bind} = -(6.3 \pm 1.0) \text{ kcal/mol}$  pari a circa  $\beta \Delta E_{bind} \simeq 10$ , che sovrastimano quanto supposto sperimentalmente e quanto ricavato da De Michele [19, 20]. Inoltre dalle loro simulazioni non è mai stato riscontrato il fenomeno della nematizzazione spontanea, nonostante avessero provato a simulare una soluzione formata da 458 segmenti di DNA, benché riuscissero ad evidenziare l'aggregazione spontanea dei filamenti in delle lunghe catene semi-rigide.

Mentre per quanto riguarda la chiralità sia levogira che destrogira esibita dal sistema, composto da soli polimeri con elicità destorsa, non è ancora possibile dare una spiegazione, qualcosa possiamo riuscire a dedurre rispetto alla dipendenza dalla tipologia di sequenza. Tutti i modelli sinora proposti sono sostanzialmente indipendenti dalla composizione chimica dei DNA in soluzione e perciò è possibile pensare di introdurre delle caratteristiche strutturali legate al tipo di sequenza presa in esame.

In questa tesi, infatti, andremo a trattare computazionalmente lo studio conformazionale di

doppie eliche di B-DNA corte in funzione della sequenza di acidi nucleici di cui sono composte. L'attenzione sarà ristretta in particolare, alle sole sequenze palindrome di 12 basi azotate (dodecameri) per cui è stata riscontrata, sperimentalmente [17], la formazione di una fase nematica in soluzione acquosa a diverse concentrazioni critiche. Proprio questo fattore ci lascia supporre che, in tale comportamento, svolga un ruolo cruciale la conformazione a differente piegatura ("bending") di determinate sequenze in accordo con quanto supposto da [17]<sup>1</sup>, e sarà esattamente grazie a questa caratteristica che le andremo a caratterizzare in maniera estensiva.

Nella prima parte del lavoro utilizzeremo la dinamica molecolare "full-atom" per simulare il sistema e questo ci consentirà di ottenere una prima stima di tale piegatura, riscontrata mediante l'introduzione di opportune grandezze osservabili ideate ad-hoc. Caratterizzeremo dapprima come il ruolo svolto dalle interazioni elettrostatiche incida sul bending dell'acido nucleico, effettuando uno studio in funzione della concentrazione di sale disciolta in soluzione per poi passare al confronto diretto della piegatura al variare delle sequenze proposte.

Il valore del bending ottenuto dalle simulazioni a singola molecola di DNA verrà poi utilizzato nella modellizzazione "coarse-grained" del doppio filamento. Nella seconda parte, quindi, introdurremo una descrizione molto semplificata della struttura del DNA piegato, assimilandolo ad un oggetto formato da un doppio cilindro sovrapposto, i cui assi sono inclinati tra loro dell'angolo sperimentalmente calcolato nella prima parte, ed aventi due siti attrattivi come già modellizzato in [19].

Infine, ricalcando l'approccio usato da De Michele [19], caratterizzeremo la termodinamica del sistema composto da tali cilindri piegati mediante l'ausilio di simulazioni Monte Carlo, che ci consentirà di esplorare l'ensemble NPT ed ottenere l'equazione di stato delle sequenze. L'analisi dell'equazione di stato corrispondente ci permetterà di risalire alla pressione a cui avviene

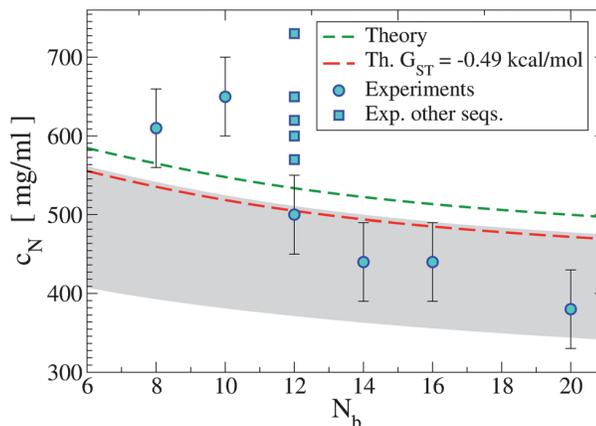


Figura 4: Diagramma estratto dal lavoro di Rovigatti e De Michele [20]. Vediamo come il modello proposto in questo articolo sia allineato all'andamento previsto per le concentrazioni di nematizzazione più basse ottenute per le sequenze di DNA in funzione del numero di basi  $N_b$ . I quadrati invece mostrano le altre concentrazioni sperimentali ottenute da Nakata et al. [16] per dodecameri di DNA fuori dall'andamento previsto.

<sup>1</sup>"We speculate that sequences whose helical conformation has a sufficient degree of bending and/or tilt of the terminal bases have an RH N0 phase [...]"

la transizione e di conseguenza stimare l'ampiezza della regione di coesistenza delle fasi. La conoscenza di tale regione, infatti, ci permette di stimare la concentrazione critica a cui il sistema nematizza in funzione della piegatura delle strutture introdotte con il nuovo modello che ipotizziamo avere un andamento non lineare a causa della particolare geometria presentata.

E' ragionevole supporre che tali oggetti piegati trovino sempre maggiore difficoltà ad ordinarsi in una fase nematica all'aumentare del grado di piegatura, spiegando quindi l'aumento della concentrazione critica riscontrato sperimentalmente [17, 16] per determinate tipologie di sequenze. A questo punto se la variazione della  $\varphi_N$  in funzione dell'angolo è compatibile con l'andamento osservato sperimentalmente il modello introdotto sarà utile a fornire una prima descrizione del fenomeno. Tale studio perciò sarà semi-quantitativo, riproducendo l'andamento previsto e consentendo una nuova interpretazione delle interazioni nei processi di "self-assembly" pur non aspettandoci di ottenere dei valori quantitativamente in accordo con quanto ottenuto sperimentalmente. Questo apre le porte a nuove sfide computazionali, in futuro il modello potrà essere ulteriormente esteso ed ottimizzato in modo da avere un migliore accordo con quanto osservato sperimentalmente, ma fornisce un nuovo strumento utile per la creazione di nuovi materiali attraverso l'utilizzo di questa nuova ed interessante proprietà, il "bending" degli acidi nucleici.

## Parte I

# Background Scientifico

# Capitolo 1

## Il ruolo del DNA nella Fisica della Materia Soffice

### 1.1 Primi passi nell'investigazione della struttura dell'acido nucleico

Prima di iniziare qualsivoglia trattazione sulla struttura conformazionale propria di una molecola di DNA, è bene conoscerne il contesto storico [22, 23], le proprietà strutturali e la nomenclatura standard adottata dai testi scientifici.

Storicamente la molecola di DNA fu isolata per la prima volta dal biochimico svizzero Friedrich Miescher nel 1869 [23], individuando nel pus di bende chirurgiche usate una particolare sostanza microscopica organica che chiamò *nucleina*, poiché rinvenuta come residuo del nucleo delle cellule di materiale organico. Successivamente a questa scoperta, nel 1878, Albrecht Kossel dapprima riuscì ad isolare la parte non proteinica della molecola, l'acido nucleico, e successivamente poté individuare le cinque principali nucleobasi da cui esso era costituito. Particolarmente importante, ai fini della comprensione della loro struttura, fu il contributo dato nel 1919 da Phoebus Levene che individuò la conformazione del nucleotide, una struttura organica composta da una base azotata, uno zucchero pentoso ed un gruppo fosfato, suggerendo che tale macromolecola fosse nient'altro che un filamento di nucleotidi legati tra loro mediante gruppi fosfati. Tale sostanza risulta perciò essere un acido, il cui nome scientifico deriva direttamente dalla composizione delle tre sub-unità fondamentali di cui è composto: l'acido desossiribonucleico (DNA).

Successivamente, con la diffusione della spettroscopia a raggi-X venne mostrato che il filamento di DNA era costituito da una struttura estremamente regolare, il che confermava, l'allora solo ipotetica, periodicità della catena molecolare. Solo nel 1944, nel libro "*What is Life?*" e dopo la grande diffusione di consensi in ambito scientifico da parte della Meccanica

Quantistica, Erwin Schrödinger ipotizzò che, poiché secondo la fisica quantistica i sistemi di pochi atomi hanno un comportamento pressoché disordinato, il materiale genetico doveva essere costituito necessariamente da una grande molecola, formata da una sequenza nucleotidica non ripetitiva, sufficientemente stabile da poter conservare l'informazione genetica che venne chiamato appunto "cristallo aperiodico".

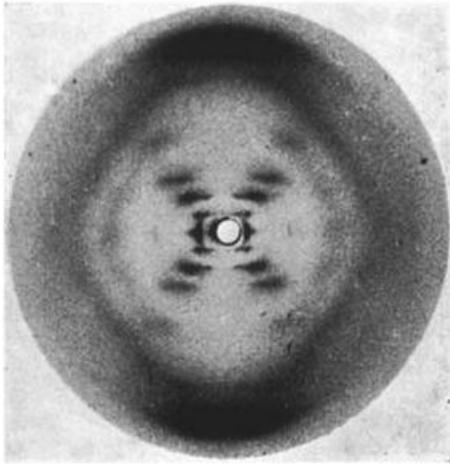


Figura 1.1: Prima immagine del DNA ottenuta tramite spettroscopia a raggi-X ad opera di Raymond Gosling e Rosalind Franklin nel 1952.

Nikolai Koltsov nel 1927 fu il primo a proporre lo schema secondo il quale i tratti ereditari degli organismi viventi venissero tramandati attraverso una grande macromolecola, "giant hereditary molecule", costituita da due filamenti complementari che potevano avere la possibilità di replicarsi. Tale ipotesi venne dapprima confermata da Frederick Griffith nel 1928 con lo studio della trasmissione genetica dei batteri *Pneumococco*, e successivamente dimostrata nel 1953 da Alfred Hershey e da Martha Chase studiando il fago T2. Essi mostrarono che il DNA posseduto dal virus iniettato all'interno della cellula batterica è responsabile della produzione di ulteriori DNA e proteine virali nei batteri *Escherichia Coli*, mentre le proteine, restando all'esterno, non sono ereditate nelle generazioni successive.

Da questo punto di vista l'acido nucleico assunse un ruolo sempre più importante in ambito biologico, e scoprirne la sua esatta struttura divenne un lavoro di primaria importanza per poterne studiare più nel dettaglio le proprietà biochimiche. L'enorme passo avanti nella comprensione della sua struttura è dovuto in gran parte agli studi di diffrazione di raggi-X effettuati da Rosalind Franklin e da Raymond Gosling nel 1952, grazie ai quali i biologi James Watson e Francis Crick nel 1953 poterono scoprire e mostrare al pubblico per la prima volta il modello, ancora oggi utilizzato, di struttura a *doppia elica* [24] dell'acido nucleico (double-helix model).

Inoltre in un'importante presentazione del 1957, Crick propose *il dogma centrale della biologia molecolare*, il principio secondo il quale il flusso dell'informazione genetica è monodirezionale e parte dagli acidi nucleici per arrivare alle proteine. In questo contesto quindi il DNA assume un ruolo centrale nella conservazione del codice genetico: l'informazione genetica è interamente conservata nel nucleo di ciascuna cellula sotto forma di una grande macromolecola organica, il DNA, che può essere replicata per poter essere utilizzata all'interno della cellula. Il trasporto dell'informazione genetica al di fuori del nucleo è effettuato mediante un altro grande polimero organico, l'RNA, strutturalmente simile all'acido desossiribonucleico, il quale viene successivamente tradotto in proteine che rappresentano la forma operativa dell'informazione

genetica.

Particolare interesse fu posto nei riguardi del meccanismo di replicazione, la cui conferma finale, basata sulla peculiare struttura a doppia elica, fu fornita nel 1958 dall'esperimento di Meselson-Stahl. In un successivo lavoro, Crick dimostrò come il codice genetico fosse basato su triplette di basi azotate non sovrapposte, permettendo ad Har Gobind Khorana, Robert Holley e Marshall Warren Nirenberg di decifrarlo. Un'altra importante scoperta fu fatta nel 1961 da Marshall Nirenberg e Severo Ochoa, che scoprirono che ogni tripletta di nucleotidi codifica uno specifico amminoacido.

Dal punto di vista termodinamico, la capacità del DNA di formare fasi liquido-cristalline quando immerso in soluzione acquosa era nota fin dagli anni '40 e giocò un ruolo cruciale nella decifrazione della sua struttura, permettendo la formazione di lunghe catene polimeriche allineate e la conseguente possibilità di studiarle mediante l'uso della spettroscopia a raggi-X. Da quel momento in poi la fase LC della soluzione di doppi filamenti di DNA è stata estensivamente caratterizzata dall'analisi ottica, dalla spettroscopia a raggi-X ed attraverso la risonanza magnetica, prendendo in esame aggregati di lunghezza variabile da  $N = 10^6$  basi azotate, tipiche di strutture polimeriche semiflessibili, a  $N = 10^2$ . Tale numero di basi è tipico di frammenti rigidi di strutture a catena, ed è comparabile con la dimensione della lunghezza di persistenza della piegatura caratteristica dell'acido nucleico, pari a  $\Lambda_p \simeq 40nm$  [26].

Attraverso questi numerosi studi riguardanti filamenti lunghi di DNA è stata portata alla luce l'esistenza di un complesso diagramma di fase svelando la presenza di differenti tipologie di aggregazione intermolecolare, sostanzialmente indipendenti dal tipo di sequenza considerata. In particolare negli esperimenti, al crescere della concentrazione di DNA in soluzione, si è riscontrata l'esistenza di una fase isotropica, una fase chirale nematica liquido-cristallina, una fase colonnare uniassiale, una fase ad alto ordine colonnare LC ed infine una fase solido-cristallina. In particolare, come mostrato dai lavori di Onseger e Bolhuis-Frenkel [14, 15], le transizioni di fase successive alla fase isotropica scomparivano quando la lunghezza dell'acido nucleico diventa comparabile al suo diametro, rendendo di fatto impossibile l'allineamento su lunga scala.



Figura 1.2: Watson e Crick presentano al mondo accademico la struttura a doppia elica del DNA.

La vera sorpresa è stata la scoperta da parte di Nakata, Zanchetta et al. [16, 17] della ricomparsa delle fasi nematiche e colonnari in catene lunghe da 6 a 20 basi azotate, contrariamente a quanto previsto dalla teoria, mediante un processo di “self-assembly” dipendente dalla tipologia di sequenza studiata: proprio in questo contesto si articola il lavoro della nostra tesi, che punterà a comprendere il ruolo della conformazione strutturale di segmenti corti di DNA in soluzione acquosa.

Tutte queste scoperte pongono i pilastri su cui poggia la moderna biologia molecolare, il cui interesse è sempre crescente e la comprensione di interessanti meccanismi legati a questa particolare macromolecola portano a traguardi scientifici di grande valore: con un curriculum di tale spessore, infatti, il DNA merita di essere studiato in qualsiasi ambito scientifico.

## 1.2 Il processo di “Self-Assembly”

Come già accennato, una delle proprietà più interessanti della materia soffice, è quella di formare aggregati mediante un processo di auto organizzazione molecolare denominato “Self-Assembly” [3]. Questo processo è dovuto alla capacità delle molecole di organizzarsi a formare grandi strutture supramolecolari senza l’intervento di forze esterne, permettendo al sistema coinvolto nella transizione di arrangiarsi in una configurazione ad alto grado di ordinamento per mezzo della minimizzazione della propria energia libera.

I soggetti coinvolti in questo processo di riorganizzazione sono molteplici ed il fenomeno si differenzia sostanzialmente in due tipologie, *intermolecolare* ed *intramolecolare*, a seconda se l’arrangiamento avviene all’interno della molecola stessa come ad esempio nel processo di “*foldng*” (ripiegamento) nel caso delle proteine in generale, chiamato anche “pairing” di singoli filamenti nel caso del DNA, oppure se tale riorganizzazione si instaura tra le molecole interagenti all’interno del sistema. Le dimensioni caratteristiche del fenomeno coinvolgono un ampio spettro di grandezze, spaziando da interazioni di piccole molecole anfifiliche, come il sapone e lunghe pochi Å, a riorganizzazioni di grandi polimeri e sostanze colloidali, le cui dimensioni tipiche arrivano anche all’ordine del  $\mu m$ .

Per quanto riguarda le interazioni intramolecolari particolarmente interessanti sono quelle responsabili del ripiegamento delle proteine su se stesse che, per esempio attraverso la scelta della composizione chimica nel sequenziamento del singolo filamento di acido nucleico, possono essere controllate per formare dei particolari strutture od addirittura degli origami [9]. Lo stesso fenomeno di “base pairing” tra coppie di basi può essere finemente controllato attraverso il sequenziamento prestabilito nei singoli “strand” di DNA per formare doppi filamenti shiftati con basi estremali spaiate, chiamate “dangling-bases”, utilizzabili a loro volta per formare legami molto stabili tra diverse doppie eliche presenti in soluzione.

Anche le forze che si instaurano tra diverse molecole responsabili del fenomeno di “self-assembly” intermolecolari spaziano tra diverse tipologie, formando legami non covalenti di

varia natura [2, 3, 4]:

- *forze di Van Der Waals*, somma di forze attrattive o repulsive di differente tipologia, come ad esempio le interazioni dipolari (interazioni di Keesom, di London o di Debye), la cui combinazione dà luogo a forze intermolecolari a lungo raggio che scalano come  $\propto 1/r^6$  e caratterizzate da energie di legame molto deboli  $E_{VDW} \simeq 0.01 \frac{eV}{molecola}$ ;
- *coordinazione tra metalli*, in cui un atomo generalmente metallico, è circondato da altre specie atomiche o ioniche;
- *interazioni  $\pi - \pi$* , chiamate anche *interazioni di  $\pi$ -stacking*, sono delle forze di natura attrattiva che si instaurano tra gli orbitali molecolari  $\pi$  di due molecole aromatiche adiacenti. Gli orbitali  $\pi$  sono formati dalla sovrapposizione laterale degli orbitali  $p_z$  degli atomi inseriti all'interno di anelli aromatici e le interazioni intermolecolari a cui danno luogo sono presenti anche nel fenomeno dell'impilamento delle basi azotate nel DNA [36];
- *legami ad idrogeno*, forze di interazione attrattiva tra molecole polari in cui l'idrogeno acquista una carica parziale positiva  $\delta^+$  quando legato ad un elemento fortemente elettronegativo, permettendogli di interagire elettrostaticamente con altre specie. Essi sono i principali responsabili del "base-pairing" cioè della formazione della coppia delle basi azotate negli acidi nucleici ed hanno energie di legame dell'ordine di  $E_{H-bond} \simeq 0.1 \div 0.5 \frac{eV}{molecola}$ ;
- *interazioni idrofobiche*, legami che si instaurano in soluzione acquosa per mezzo dell'esclusione delle molecole d'acqua. Una sostanza non polare si dissolve difficilmente in soluzione poiché non può partecipare alla formazione di legami ad idrogeno con il solvente. Quando viene immersa in soluzione, l' $H_2O$  circonda la molecola ingabbiandola. Per preservare la struttura tetraedrica del solvente, le molecole d'acqua che formano la "gabbia" si devono orientare in modo da evitare la formazione di legami-H con la molecola e questa riorganizzazione molecolare, molto sfavorevole dal punto di vista entropico, rende energeticamente dispendioso inserire la molecola polare all'interno del solvente. L'effetto di questo svantaggio entropico è quello di creare una forza attrattiva tra due molecole polari immerse in soluzione acquosa, di modo tale che le due lascino scoperta la minore superficie possibile alla gabbia di  $H_2O$  che le circonda. Tale effetto combinato ad altre interazioni, come le aromatiche appena descritte, è responsabile dell'impilamento delle basi azotate nei filamenti di acidi nucleici, dando luogo anche ad interazioni intermolecolari responsabili dello stacking "end-to-end" [16, 36];

- *interazioni elettrostatiche od elettromagnetiche*, dovute alla presenza di dipoli indotti o di fenomeni di polarizzabilità oppure in generale legate a tutti quei fenomeni di ricombinazione di carica all'interno di molecole.

Tutte queste interazioni, che spesso si trovano combinate tra loro a formare una grande complessità di legami ed aggregazioni, concorrono insieme nel mettere in atto il processo di “self-assembly” intermolecolare, il cui risultato principale è l'estensione del grado di ordinamento locale a lunghezze caratteristiche molto superiori a quelle tipiche delle dimensioni atomiche, rendendo l'effetto di fatto macroscopicamente visibile e controllabile attraverso variabili macroscopiche esterne.

Come abbiamo avuto modo di accennare in precedenza le proprietà di aggregazione degli acidi nucleici sono vitali nelle funzioni basilari della cellula biologica e sono state estensivamente studiate nel campo delle biotecnologie. Il processo di “self-assembly” di un singolo filamento di ssDNA (“single-stranded” DNA) nella caratteristica forma a doppia elica, è un processo di centrale importanza nella biotecnologia: viene usato soprattutto come impalcatura nella rivelazione del DNA, nell'assemblamento programmabile di nanostrutture, nel trasporto direzionale di carica e nei processi di nanofabbricazione. Un altro fenomeno di eccezionale importanza è la condensazione del DNA, dove i controioni trasformano la repulsione elettrostatica tra i DNA spogli in una attrazione reciproca, facilitando il packaging di DNA a doppio filamento nei nuclei delle cellule e nelle capsidi virali.

Recentemente è stata scoperta una nuova via di aggregazione tra gli acidi nucleici, cioè l'aggregazione “end-to-end” spontanea di doppie eliche di DNA corte in una struttura ad asta flessibile [16] osservata mediante l'evaporazione di acqua da una soluzione ad alta concentrazione di DNA: nella struttura residua viene osservata una fase liquido cristallina. Poiché le dimensioni dei frammenti presi in esame sono comparabili in lunghezza e larghezza, l'osservazione di tale ordinamento assiale in catene rettilinee può essere spiegato solamente mediante l'aggregazione in una struttura supramolecolare, suggerendoci quindi un'attrazione “end-to-end” tra le basi di ciascuna delle doppie eliche, ad opera di un processo di auto aggregazione spontanea.

Gli “short duplex” di DNA, mediante le forze di “stacking” coassiale tra basi terminali favorite da interazioni idrofobiche, possono quindi auto assemblarsi a formare catene debolmente legate [16, 17]. Tale processo di polimerizzazione reversibile è sufficiente a promuovere il mutuo allineamento di queste catene e la successiva formazione di un ordinamento macroscopico orientazionale, rendendo così nuovamente possibile la formazione di una fase nematica liquido-cristallina (LC).

Come già accennato la capacità delle lunghe molecole di acido nucleico di formare fasi liquido-cristalline era già nota agli studiosi fin dagli anni '40, quando grazie alla spettroscopica a raggi-X fu possibile studiarne in modo dettagliato la struttura. Da questo punto di

vista, però, la possibilità di un ordinamento colonnare non era sorprendente, poiché venivano considerati filamenti estremamente lunghi costituiti da un numero di basi azotate superiore alle centinaia di unità: una macropolimero del genere, estremamente rettilineo e rigido, trova molto conveniente formare una fase liquido-cristallina, non mostrando evidenze sperimentali di un eventuale processo di “self-assembly” mediato da forze di “stacking”. La vera sorpresa è stata ritrovare una fase LC anche per frammenti molto corti formati da  $N < 20$  coppie [17, 16], chiaro indicatore della presenza di forze di attrazione “end-to-end” tra basi terminali che acquistano sempre maggiore rilevanza al diminuire della lunghezza del segmento di DNA.

Alla luce di ciò, il processo di auto-aggregazione è uno dei principali metodi per la realizzazione di materiali con caratteristiche innovative, grazie all’enorme quantità di parametri controllabili dati dalla scelta delle caratteristiche dei singoli costituenti. In particolare è possibile controllarne la forma, per esempio decidendone a priori la geometria oppure introducendo eventuali asimmetrie nella struttura (“base dangling”), il grado di flessibilità o di rigidità, la valenza intesa come la capacità di aggregarsi con un differente numero di altri costituenti oppure decidere a priori la tipologia di interazione tra i blocchi fondamentali.

Queste numerose proprietà, inoltre, possono essere sfruttate per replicare a diverse scale di lunghezza e di tempo il comportamento del sistema atomico: la sfida tecnologica dietro lo studio della materia soffice è perciò quella di cercare di poter utilizzare un così vasto set di interazioni conosciute per la realizzazione di nuovi materiali con caratteristiche innovative. In particolare una comprensione più approfondita del processo fisico alla base delle interazioni di “stacking” è di grande interesse scientifico non solo perché possono essere utilizzate dai ricercatori per la costruzione di grandi aggregati molecolari, ma anche perché esse vengono spesso considerate come la via prebiotica per spiegare il passaggio dalla sintesi casuale di molecole elementari a base di carbonio alla formazione delle prime molecole organiche complesse <sup>1</sup> in grado di effettuare la sintesi proteica. Mentre è stato studiato dettagliatamente, sia per via sperimentale che teorica, l’effetto di forze “side-by-side” tra lunghe molecole di DNA, poco ancora si conosce al riguardo delle condizioni del meccanismo microscopico dell’associazione “end-to-end” ed inoltre tale effetto non è stato mai investigato nei processi biologici e tecnologici.

In questo contesto risulta davvero molto interessante l’investigazione numerica del ruolo giocato dalla tipologia di sequenza di basi nella determinazione della conformazione del duplex di DNA. Questo studio porterebbe utili informazioni per mettere alla prova l’approccio teorico sviluppato da De Michele [20, 19] e ci consentirebbe di sviluppare un modello minimalista “coarse-grained” a dDNA piegato, modellizzando la struttura con un doppio cilindro piegato con due siti di interazione che mediano le forze idrofobiche di “stacking”.

---

<sup>1</sup>Tali molecole primordiali erano probabilmente oligomeri di RNA

## Capitolo 2

# Struttura del DNA

### 2.1 Composizione Chimica

Gli *acidi nucleici* sono delle macromolecole aperiodiche organiche presenti in tutti i sistemi biologici, all'interno dei quali svolgono il ruolo di trasporto e conservazione dell'informazione genetica. A seconda della composizione chimica e della struttura da cui sono costituiti, vengono classificati in due tipologie distinte: l'RNA, ossia l'acido ribonucleico, che si trova nei sistemi biologici sotto forma di singolo filamento, ed il DNA, che andremo ad analizzare in maggior dettaglio.

In natura, il DNA è presente sia sotto forma di singolo (sDNA) sia come doppio (dDNA) filamento, avvitato attorno ad un asse a formare una struttura elicoidale: anche se in realtà esistono forme ben più complesse a tre e quattro filamenti, in questa sede siamo interessati alla sua forma più classica, quella a doppia elica.

L'acido desossiribonucleico (DNA - Deoxyribo Nucleic Acid) è un lungo polimero organico costituito da monomeri, ovvero da unità fondamentali ripetute disposte a formare una catena, chiamate nucleotidi (o desossiribonucleotidi<sup>1</sup>, quando vi è ambiguità di classificazione con i ribonucleotidi presenti nell'RNA). Tali nucleotidi sono esteri fosforici dei nucleosidi, cioè composti organici prodotti dalla reazione di acetilazione di un alcol o di un fenolo con acido carbossilico o un suo derivato, che a loro volta sono altri composti organici costituiti da uno zucchero

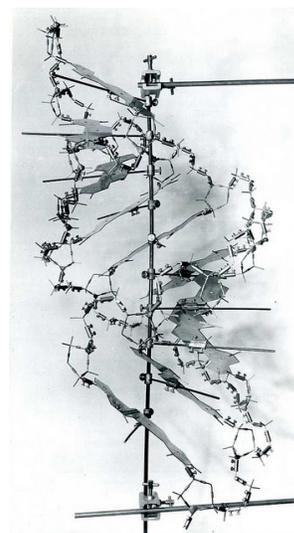


Figura 2.1: Modello di DNA presentato da Watson-Crick.

<sup>1</sup>In questa sede ometteremo sempre il prefisso desossi-, od equivalentemente deossi-, poiché tratteremo esclusivamente il DNA.

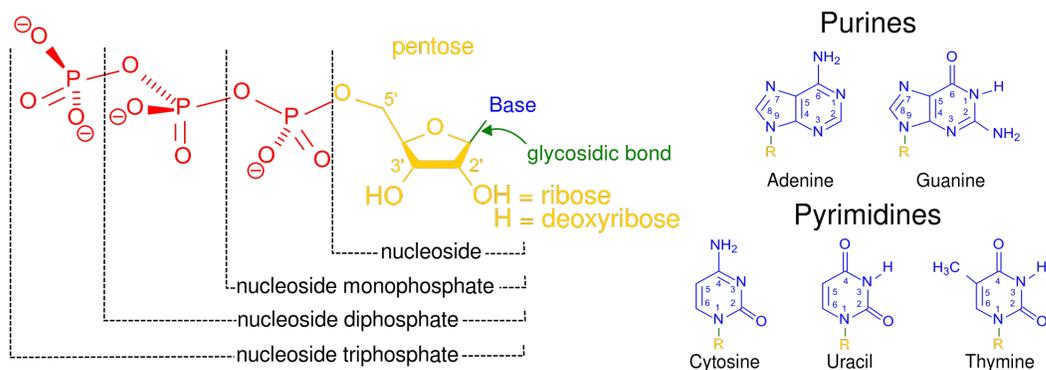


Figura 2.2: Struttura schematica di un intero nucleotide. A destra troviamo la composizione strutturale delle differenti basi azotate che andranno a costituire la sequenza. L'uracile è una base azotata che troviamo solamente nell'RNA al posto della timina.

pentoso (D-ribosio o 2-desossi-D-ribosio) legato ad una base azotata.

La struttura di un nucleotide perciò è costituita da tre sub-unità principali connesse tra loro, la cui ripetizione lungo l'asse del principale del DNA e l'appaiamento planare perpendicolarmente a tale asse, ne caratterizzerà la sequenza:

- una base azotata, di tipologia purinica o timinica;
- uno zucchero pentoso, un monosaccaride contenente 5 atomi di carbonio;
- un gruppo fosfato, costituito dall'anione fosfato legato ad una molecola organica tramite 4 ossigeni;

E' impossibile parlare di composizione chimica del DNA nel senso stretto del termine infatti, essendo un polimero aperiodico, possiede un numero arbitrario di nucleotidi, i quali composti a loro volta da diverse basi azotate. Tale aperiodicità rende la composizione chimica, come pure la massa del biopolimero, strettamente dipendente dal tipo di sequenza considerata. In questo caso perciò, si parla di *struttura primaria* [53] dell'acido nucleico, in riferimento alla composizione atomica della sequenza di basi azotate, di zuccheri pentosi e di gruppi fosfati, con cui è costituita la macromolecola in oggetto. Poiché in questo testo ci riferiremo esclusivamente ad una tipologia di DNA, potremo identificarlo senza ambiguità mediante la trascrittura della sequenza di basi di cui esso è costituito (ad esempio AATGAATTCATT), usando per brevità la notazione adottata in [17] nella quale vengono riportate esclusivamente le prime tre basi (perciò AAT).

Quindi per poter effettuare uno studio più dettagliato, è necessario analizzare ciascuna sub-unità monomerica dal punto di vista della composizione chimica, dalla disposizione geometrica degli atomi nella stessa e dei legami che intercorrono per formare tale struttura elicoidale.

Nell'assemblamento delle basi non è implicitamente codificata la disposizione ad elica che esse assumeranno una volta costituita la macromolecola: tale fenomeno è solo una conseguenza secondaria dell'impilamento, lo "stacking", delle basi azotate, che vederemo avere una struttura sostanzialmente planare.

### 2.1.1 Basi Azotate

Come abbiamo già detto in precedenza, la struttura di ciascun nucleotide è caratterizzata principalmente da una sotto-unità legata ad una struttura esterna, chiamata base azotata, che rende la composizione chimica dell'acido nucleico dipendente dal tipo di sequenza considerata.

Le basi azotate, componenti chiave della codifica genetica dell'informazione, possono essere di due tipologie, a seconda della molecola di base da cui derivano:

1. *Puriniche*, cioè derivanti dalla struttura della Purina: *Adenina* e *Guanina*;
2. *Pirimidiniche*, cioè che derivano dalla Pirimidina: *Citosina*, *Timina* e *Uracile*<sup>2</sup>.

Le basi di tipo purinico sono molecole *aromatiche* formate da due anelli eterociclici pentagonali ed esagonali azotati (cioè anelli composti non solamente da atomi di carbonio) derivati dalla purina. La loro struttura ricorda la giunzione tra un poligono esagonale ed uno pentagonale disposti all'incirca su di uno stesso piano. Alla classificazione di Purine, appartengono due basi azotate: l'*Adenina* (A), con formula chimica  $C_5H_5N_5$ , e la *Guanina* (G), strutturalmente simile all'adenina ma con un ossigeno in più a formare un legame chimico covalente doppio con un carbonio nell'anello aromatico esagonale  $C_5H_5N_5O$ .

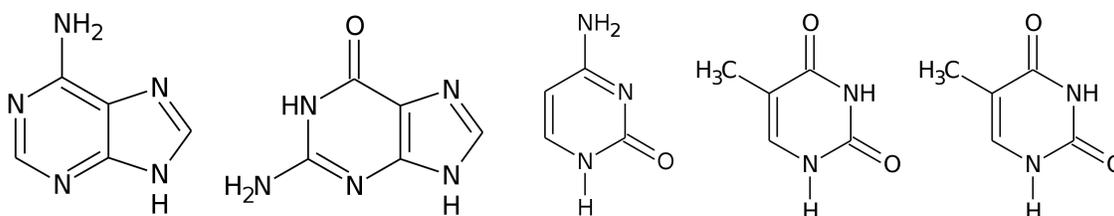


Figura 2.3: Formula strutturale delle cinque basi azotate: Adenina e Guanina, Citosina, Timina e Uracile.

Le basi pirimidiniche sono invece caratterizzate da un singolo anello aromatico esagonale eterociclico, tipico della pirimidina  $C_4H_4N_2$ , che opportunamente modificato da luogo alla formazione delle tre basi azotate anch'esse a struttura planare: *Timina* (T), di formula chimica  $C_5H_6N_2O_2$ , *Citosina* (C) la cui formula è  $C_4H_5N_3O$  ed infine l'*Uracile* (U)  $C_4H_4N_2O_2$  che troviamo esclusivamente nell'RNA in sostituzione della Timina e che perciò non andremo ad analizzare estensivamente.

<sup>2</sup>L'uracile è presente esclusivamente nell'acido ribonucleico, l'RNA.

Per differenziare i diversi atomi presenti nelle strutture aromatiche delle basi azotate, si introduce una nomenclatura particolare: partendo da uno degli atomi di azoto presenti nelle strutture esagonali, chiamato *N1*, ciclicamente si assegna un numero crescente ad ogni atomo della base considerata. I codici con cui vengono identificati gli atomi nelle basi sono riportati nelle figure 2.7 e 2.8.

Le basi azotate appena descritte costituiscono il cuore della doppia elica di DNA, appaiandosi tra loro e impilandosi a coppie l'una sull'altra. Proprio l'impilamento delle coppie nucleotidiche porterà alla formazione della caratteristica struttura elicoidale nella quale le basi azotate si disporranno in posizione centrale sostenute da una struttura esterna chiamata "backbone".

### 2.1.2 Zucchero Pentoso

Legata a ciascuna base azotata troviamo un monosaccaride, cioè un composto organico formato da carbonio, ossigeno ed idrogeno, chiamato zucchero pentoso poiché caratterizzato dalla presenza di cinque atomi di carbonio disposti su una geometria pentagonale. Nel DNA tale zucchero è il desossiribosio, con formula chimica  $C_5H_{10}O_4$ , così chiamato perché nella composizione chimica è presente un atomo di ossigeno in meno rispetto al ribosio,  $C_5H_{10}O_5$ , che viene invece impiegato nella costruzione dell'RNA.

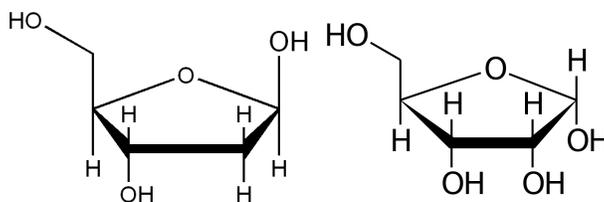


Figura 2.4: Le due tipologie di zuccheri pentosi riscontrate negli acidi nucleici, a sinistra il desossiribosio, a destra il ribosio. Notiamo in particolare l'assenza di un ossigeno nella struttura del desossiribosio.

Attraverso una reazione di idrolisi, cioè una condensazione inversa con eliminazione di acqua, lo zucchero pentoso si lega ad una base azotata formando una struttura denominata nucleoside nel caso del legame col ribosio, o desossi-nucleoside quando lo zucchero è il desossiribosio (chiamato per brevità anche nucleoside, quando non ci sono ambiguità di riferimento all'RNA).

I composti formati da base azotata e zucchero pentoso vengono così classificati, a seconda della base con cui legano:

- (desossi)-Ribosio + Adenina = (desossi)-Adenosina;
- (desossi)-Ribosio + Guanina = (desossi)-Guanosina;

- (desossi)-Ribosio + Timina = (desossi)-Timidina;
- (desossi)-Ribosio + Uracile = (desossi)-Uridina;
- (desossi)-Ribosio + Citosina = (desossi)-Citidina;

Anche lo zucchero pentoso, data la disposizione pentagonale degli atomi di carbonio, sembrerebbe essere caratterizzato da una geometria planare, in realtà però non è così ed inoltre quando si lega alla base azotata formando il nucleoside, tale struttura pentagonale non si allinea nemmeno al piano contenente la base azotata. Di questo fatto ne dovremo tenere conto in seguito quando parleremo della struttura del polimero.

In questo caso la nomenclatura proposta..

### 2.1.3 Gruppo Fosfato

Il gruppo fosfato di fatto costituisce il gruppo funzionale dell'acido nucleico. Esso ha una struttura ben precisa e conferisce la medesima reattività chimica ai composti che contengono lo stesso gruppo, è perciò centro della reattività chimica della molecola e conferisce le proprietà di acidità dell'acido nucleico.

In particolare il gruppo fosfato è un composto acido fortemente polare, costituito dall'anione fosfato  $PO_4^{3-}$  e legato alla molecola organica (nel nostro caso al nucleoside) tramite uno dei suoi 4 ossigeni. La massa molare di tale composto è pari a  $94.97g/mol$  e la sua presenza nella struttura del dDNA determina una carica netta negativa totale pari a:

$$Q_{DNA}^{tot} = -s(N - 1)e \quad (2.1)$$

dove ovviamente  $e = 1.601 \cdot 10^{-19}C$  è la carica dell'elettrone,  $N$  indica il numero di basi di cui è costituito il singolo filamento ed  $s$  è il fattore che indica il numero di filamenti dell'acido nucleico, nel nostro caso quindi  $s = 2$ . Il contributo di riscaldamento del numero di basi è dovuto al fatto che il gruppo fosfato connette due nucleosidi successivi, perciò per ogni filamento ce ne saranno  $N - 1$ , salvo diversamente indicato<sup>3</sup>.

In generale il nome del gruppo viene espresso dalla formula chimica  $(R - OPO_3)^{2-}$  la cui carica negativa si trova formalmente sul fosforo, mentre in realtà è in risonanza con tutti gli atomi della molecola.

Nei legami formati all'interno del DNA ed in condizioni di debole acidità viene trovato nella sua forma più comune, il diidrogeno-fosfato  $(H_3PO_4)^-$  il quale, come detto in precedenza determinare a, conferisce all'acido nucleico una carica elettrica negativa.

<sup>3</sup>Per esempio tale formula non varrebbe nel caso di DNA fosforilato, in cui vengono aggiunti due gruppi fosforici alle basi terminali, in questo caso la carica sarebbe perciò pari a  $Q_{DNA}^{tot} = -sNe$ .

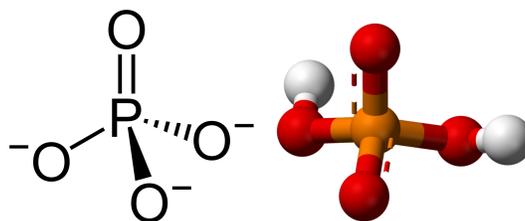


Figura 2.5: A sinistra la rappresentazione del gruppo fosfato generico, a destra il gruppo diidrogenofosfato presente nel DNA. Notiamo come un ossigeno abbia una carica elettrica scoperta, anche se come detto in precedenza tale carica viene rappresentata come formalmente appartenente al fosforo.

Il gruppo fosforico può legare uno dei suoi ossigeni, attraverso un legame fosfodiesterico di tipo covalente, al pentosio in posizione 3' oppure in 5' formando una struttura chiamata 3'-oppure 5'-nucleotide, a seconda della posizione in cui avviene il legame. Sarà proprio questa asimmetria nel legame fosforico tra nucleosidi successivi a determinare l'avvitamento della struttura elicoidale negli acidi nucleici.

## 2.2 L'accoppiamento delle basi azotate

La *struttura secondaria* degli acidi nucleici prende in esame le interazioni e gli eventuali accoppiamenti tra basi azotate, e può essere rappresentata mediante la lista delle coppie di basi che costituiscono la molecola [28].

Solitamente il DNA esiste in natura sotto forma di doppia elica, caratterizzata da una sequenza di basi che costituiscono il primo filamento, ciascuna delle quali accoppiate alle corrispondenti del secondo. Tale conformazione non è univoca, infatti la doppia elica può esistere anche se tutte le sue basi non sono accoppiate, ad esempio formando delle strutture in cui le ultime coppie di ciascuna elica sono libere, oppure formando complessi accoppiamenti a tre o quattro basi. Ovviamente in questo contesto ci riferiremo esclusivamente al primo caso.

La biologia molecolare ha messo in evidenza che le interazioni responsabili dell'accoppiamento delle basi azotate sono i legami ad idrogeno che si instaurano tra esse. Chimicamente tali legami sono favoriti tra basi azotate appartenenti a famiglie diverse: una base purinica si legherà ad una pirimidinica perciò, secondo lo schema proposto da Watson-Crick, l'adenina si legherà esclusivamente alla timina mentre la guanina alla citosina. Anche in questo caso le basi hanno diversi modi per organizzarsi, ad esempio l'accoppiamento A-T e G-C è possibile anche secondo il modello di Hoogsteen riportato in 2.6, ma le strutture che andremo ad analizzare saranno esclusivamente di tipo W-C.

Tale coppia viene chiamata “base pair” (per brevità bp) ed il relativo processo di accoppiamento tra basi, ottenuto mediante diversi legami ad idrogeno, “base-pairing”.

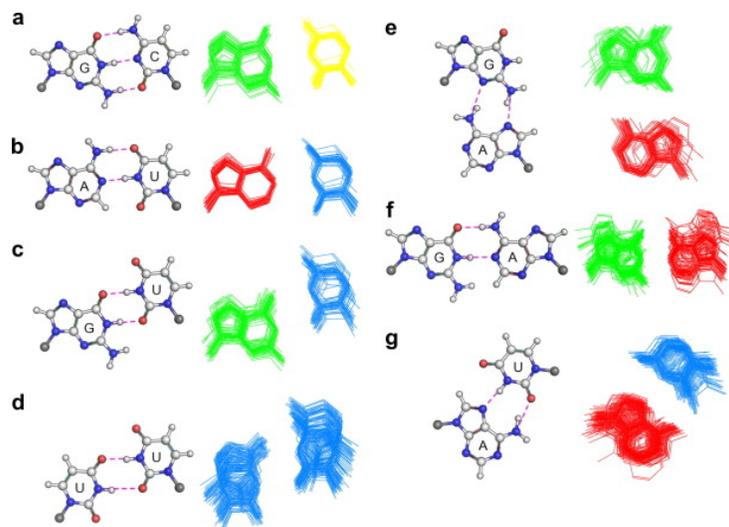


Figura 2.6: Esempio di due differenti schemi di “base-pairing” tra le diverse basi azotate: a-b) il classico schema di accoppiamento proposto da Watson-Crick nei casi  $G - C$   $A - U$ , c-d) gli accoppiamenti secondo il modello “wobble” di  $G - U$  e  $U - U$ , e) “sheared-pairing” tra  $G - A$ , f) “reversed Watson-Crick pairing”  $G - A$ , modello complementare al classico ed infine il g) modello di accoppiamento secondo lo schema di Hoogsteen tra  $A - U$ . In questa situazione [38] sono rappresentati alcuni fenomeni di “base-coupling” tra basi dell’RNA, ma la situazione è del tutto analoga nel DNA con la sostituzione  $U \rightarrow T$ .

Il legame ad idrogeno è un particolare tipo di legame determinato da una interazione elettrostatica in cui un atomo di idrogeno si lega ad un atomo fortemente elettronegativo, quale ad esempio il carbonio, il fluoro o l’azoto. Data la sua natura prettamente elettrostatica, è considerato più una forte interazione attrattiva dipolo-dipolo piuttosto che un legame vero e proprio e tale interazione, e poiché è caratterizzato da energie dell’ordine di  $5 \div 30 kJ/mole$ , risulta più debole rispetto ai legami covalenti o ionici ma più forte delle interazioni di Van Der Walls.

Gli accoppiamenti energeticamente favorevoli sono quelli tra basi appartenenti a famiglie diverse, a causa delle conformazioni strutturali proprie delle due tipologie: dato che le piramidine sono molto grandi, la coppia formata tra loro avrebbe le basi troppo distanti per poter essere stabilizzata da legami ad idrogeno, al contrario la coppia purinica le avrebbe troppo vicine e vincerebbero le forze di repulsione.

Gli unici legami interessanti perciò sono quelli purino-pirimidinici A-T e C-G: Il primo viene stabilizzato attraverso due legami ad idrogeno, mostrati in figura 2.7, mentre il secondo da tre, in figura 2.8, risultando quindi più stabile del precedente. La stabilità della doppia elica, quindi, è anche stimabile dalla quantità di coppie C-G presenti, rispetto alle più debolmente legate A-T.

In particolare la coppia di basi azotate si dispone in una geometria pressoché planare:

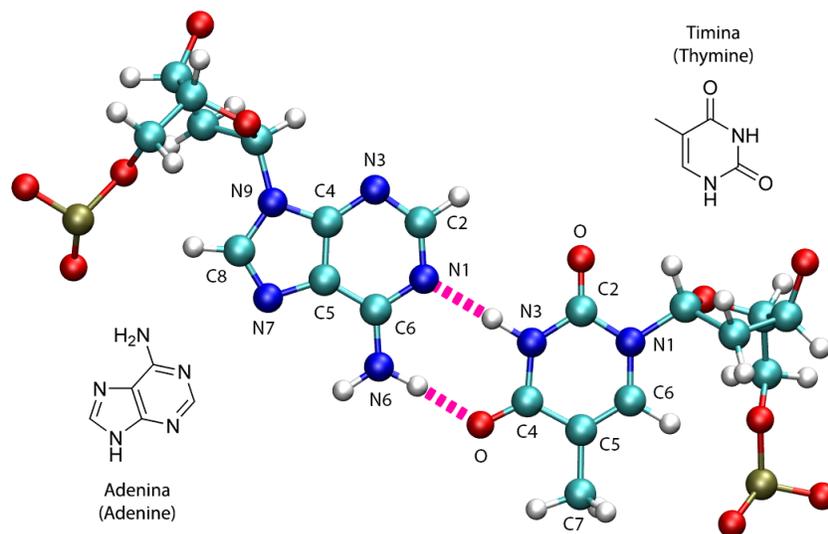


Figura 2.7: Formula strutturale e composizione chimica dell'adenina (A) e della timina (T). Le connessioni tratteggiate H-N1 ed H-O mettono in evidenza i due legami ad idrogeno che si instaurano tra le basi, permettendo la stabilizzazione della coppia. Notiamo inoltre la struttura sostanzialmente planare che esse formano.

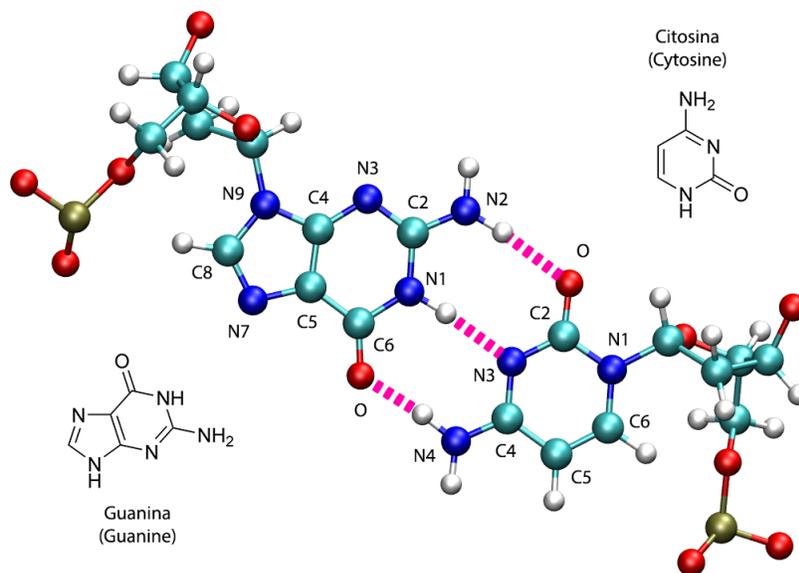


Figura 2.8: Formula strutturale e composizione chimica della guanina (G) e della Citosina (C). Stavolta i legami ad idrogeno sono tre (H-O, H-N3, H-O) e consentono la formazione di una coppia più stabile della precedente.

tutti gli anelli aromatici pirimidinici e purinici come pure i loro rispettivi legami ad idrogeno, giacciono quindi sullo stesso piano, il quale non comprende però lo zucchero pentoso e il gruppo fosfato.

Le basi una volta accoppiate<sup>4</sup>, trovano favorevole dal punto di vista energetico impilarsi parallelamente tra loro mediante le cosiddette forze di “*stacking*” idrofobico: infatti tali molecole non polari preferiscono esporre la minor superficie possibile a contatto con le molecole d’acqua disponendosi l’una sopra l’altra. Inoltre all’interno degli anelli aromatici la posizione degli orbitali  $p_z$ , perpendicolari all’anello, favorisce l’interazione con quelli della base successiva, formando un legame intermolecolare non covalente di tipo  $\pi - \pi$  [29], chiamato anche *interazione aromatica*. Tale sovrapposizione delle coppie (ma anche della singola base azotata negli acidi nucleici a singolo filamento) viene ulteriormente stabilizzata nella struttura del nucleotide dal gruppo fosforico, che lega due zuccheri pentosi successivi di due nucleosidi successivi.

## 2.3 Morfologia Strutturale della Doppia Elica

Quando ci si riferisce all’esatta geometria di posizionamento degli atomi all’interno di un acido nucleico, si parla di *struttura terziaria* del DNA.

Come già accennato dal punto di vista morfologico la molecola del DNA si può organizzare, impilando le basi azotate appaiate mediante legami ad idrogeno, in una struttura elicoidale a doppio filamento (“duplex”, denotato spesso per brevità con la sigla dDNA) avvitata attorno ad un asse principale. In realtà è possibile trovare anche sequenze di DNA a singolo filamento (“single strand”) e ben più complesse strutture a tre e quattro filamenti (“triplex” e “quadruplex”), anche se in questa sede non siamo interessati a tali aggregati seppur molto interessanti dal punto di vista di fenomeni di auto assemblamento [7, 8].

Uno dei principali responsabili della formazione della struttura elicoidale è il legame fosfodiesterico che intercorre tra un atomo di ossigeno del gruppo fosforico ed il carbonio in posizione 3’ dello zucchero pentoso di un nucleotide, con lo stesso legame dello stesso gruppo fosforico ma stavolta con il carbonio in posizione 5’ del nucleotide precedente. Questa anisotropia geometrica di legame, unita alla disposizione parallela delle basi a mezzo delle forze di stacking, comporta una differente angolazione nel posizionamento relativo tra le due coppie di nucleotidi consecutivi, permettendo al doppio filamento di avvitarsi su se stesso. Il DNA, od anche più in generale tutti gli acidi nucleici, sono dunque dei polinucleotidi in cui il gruppo fosforico fa da ponte tra le posizioni 3’ e 5’ di due unità di ribosio consecutive, le cui coppie di basi azotate sono disposte parallelamente grazie all’intervento delle forze di stacking.

E’ possibile trovare il doppio filamento di DNA in un gran numero di diverse conformazioni, denominate usando alcune lettere dell’alfabeto (usando le appendici A-, B-, C-, D-, E-, H-, L-,

---

<sup>4</sup>In realtà questo avviene anche quando le basi non si accoppiano tra loro, per esempio nel singolo filamento di DNA oppure nell’RNA.

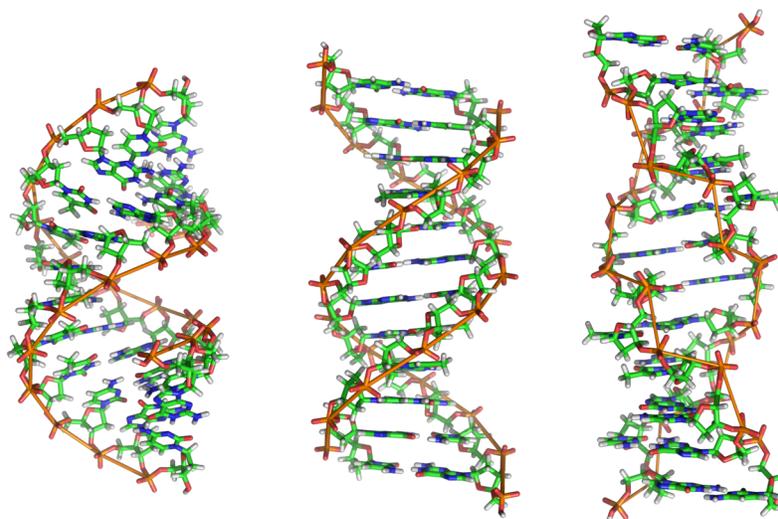


Figura 2.9: Differenti tipologie di forma per la struttura del DNA, nell'ordine A- B- e Z-. In questa illustrazione sono rappresentate solamente le più comuni ed anche le uniche strutture riscontrate nei sistemi biologici.

P-, Z-DNA), anche se in natura sono state osservate in sistemi biologici solamente le strutture A, B e Z. Nonostante tali strutture si differenziano tra loro per la tipologia di avvitamento (ad esempio la Z ha un caratteristico avvitamento levogiro a “zig-zag”) e dall'inclinazione delle basi rispetto all'asse principale, noi siamo interessati allo studio conformazionale della più frequente di tali tipologie, cioè la forma B. D'ora in avanti quando parleremo di DNA, per non appesantire troppo la notazione, ci riferiremo perciò alla struttura standard a doppio filamento del B-DNA, salvo differenti precisazioni.

La doppia elica del B-DNA è una spirale destorsa che, con l'avvitarsi dei due filamenti attorno all'asse principale, lascia esposti dei solchi di diversa larghezza. Il primo viene chiamato *solco* oppure *avvitamento maggiore* (“major groove”) ed è largo  $2.2nm$  mentre il più piccolo, *solco* od *avvitamento minore* (“minor groove”), è largo circa la metà del precedente, cioè circa  $1.2nm$ . Questo fatto implica, tra le altre cose, una maggiore accessibilità, cioè una maggiore facilità ad interagire con altri sistemi, a quelle basi che si trovano nell'avvitamento maggiore contrariamente a quanto accade nell'avvitamento minore, in cui esse sono più protette dalla struttura esterna compatta.

In condizioni di rilassamento, la struttura principale del DNA ha geometria grosso modo approssimabile ad una forma cilindrica caratterizzata dalla concatenazione e dall'impilamento delle basi azotate, che si dispongono a  $d_{bp} = 0.34nm$  l'una sull'altra, mentre in generale la larghezza varia tra  $2.2 \div 2.6nm$ . Per i nostri scopi possiamo considerare il raggio medio della struttura cilindrica pari a  $R = 1.2nm$  e grazie a queste informazioni possiamo facilmente calcolare il volume del cilindro che racchiude la struttura elicoidale, definendo prima la lunghezza

Geometria	A-DNA	B-DNA	Z-DNA
Senso di avvvitamento dell'elica	destrogiro	destrogiro	levogiro
Forma e dimensione	larga e corta	intermedia	lunga e stretta
Solco maggiore	-	2.2nm	-
Solco minore	-	1.2nm	-
Unità strutturali ripetute	1bp	1bp	2bp
Rotazione/bp ( $\theta$ )	33.6°	35.9°	60°/2
Rotazione completa ogni $N_{rot}$ bp	11	10.5	12
Inclinazione ( $\phi$ ) di ciascuna bp	+19°	-1.2°	-9°
Distanza di impilamento ( $d_{bp}$ )	0.24nm	0.34nm	0.37nm
Passo dell'elica	2.53nm	3.54nm	4.56
Diametro ( $D = 2R$ )	2.55nm	2.37nm	1.84nm

Tabella 2.1: Tabella riassuntiva delle caratteristiche generali della doppia elica nelle tre differenti tipologie di DNA presenti in sistemi biologici.

totale della sequenza:

$$L_{DNA} = N d_{bp} \quad (2.2)$$

e successivamente calcolando il volume approssimato della sua struttura:

$$V_{DNA} = \pi R^2 L_{DNA}. \quad (2.3)$$

E' altresì utile conoscere che il vettore perpendicolare a ciascuna base azotata è inclinato rispetto all'asse principale di  $\phi = -1.2^\circ$  mentre la rotazione planare di impilamento, che determina l'elicità della struttura, è di  $\theta = 35.9^\circ$  di una coppia di basi rispetto alla precedente. Questo implica che il B-DNA compie una rotazione completa attorno all'asse principale ogni  $N_{rot} = 10.5$  coppie di basi.

## 2.4 Nomenclatura: il formato \*.pdb

La collocazione spaziale degli atomi in una proteina viene descritto mediante un "database" contenente tutte le informazioni utili a decifrarne la sua struttura terziaria. Esistono molti tipi di formati per salvare la geometria della molecola in un file, ma in questa sede andremo ad utilizzare esclusivamente il formato con il quale sono salvate le configurazioni nell'archivio *Protein Data Bank*: il formato \*.pdb.

Convenzionalmente è un file di testo in cui nell'intestazione (HEADER) vengono salvati dati riguardanti il titolo (TITLE), nome dell'autore (AUTHOR), alcune informazioni descrittive generiche (REMARK) e la sequenza dei residui presenti nella proteina (SEQRES).

La parte più interessante è però quella che descrive i singoli atomi (ATOM) di cui è costituita la molecola, nella quale troviamo diverse colonne: Nella prima vengono enumerati

gli atomi, nella seconda viene descritto il tipo di atomo usando le convenzioni di nomenclatura descritte in precedenza, nella terza troviamo specificata la tipologia di residuo in cui è inserita (nel nostro caso A, T, C, G) mentre la quarta indica a quale filamento appartengono. Le posizioni cartesiane degli atomi sono contenute nelle tre colonne successive e descritte da tre valori “float” espressi in Angstrom ( $\text{\AA}$ ) a cinque cifre significative, anche se in realtà è possibile aumentare la precisione della descrizione geometrica usando dei numeri “double” a precisione doppia. Le ultime tre colonne sono il numero di occupazione, il fattore di temperatura e il nome chimico generico dell’elemento.

Grazie all’ausilio di questo formato saremo in grado di esaminare nel dettaglio la dinamica della proteina in soluzione e di poterne effettuare l’analisi delle sue caratteristiche geometriche e conformazionali.

## Capitolo 3

# Simulazione e Modellizzazione Computazionale del DNA

Lo stato dell'arte per quanto concerne la simulazione del comportamento di grandi polimeri organici sta evolvendo davvero rapidamente, stimolato dai progressi nei campi della nanotecnologia [30] e dalla sempre maggiore potenza di calcolo ed efficienza computazionale dei supercalcolatori a disposizione dei ricercatori. Molteplici modellizzazioni di differenti tipologie di acidi nucleici sono disponibili in letteratura ed il loro utilizzo dipende in modo cruciale da quale caratteristica si vuole simulare nel sistema: partendo dalla completa rappresentazione atomica, nella quale tutti gli atomi, solventi e ioni inclusi, sono considerati esplicitamente in un database che descrive con precisione la struttura terziaria della molecola (una tra le più importanti risorse è la Protein Data Bank), fino ad arrivare ad un modello “coarse-grained” in cui gruppi di atomi sono modellizzati mediante la costruzione di unità fondamentali (“beads”, perline) connesse tra loro a formare una catena vermiforme [20] aventi delle determinate proprietà di interazione tra i suoi costituenti. In alcuni casi particolari, determinate proprietà possono essere investigate con un approccio ancora più semplificato [15], considerando il polimero come approssimabile ad una superquadrica di forma cilindrica, come fatto in [19].

L'approccio “full-atom” implementato nella dinamica molecolare mediante l'utilizzo di “force-field” (spesso abbreviato con F-F), come per esempio AMBER [32] o CHARMM [33], è il modo più dettagliato per investigare la struttura interna delle molecole di DNA [34]. In questo contesto i “force-field” regolano l'interazione atomica nel sistema che viene simulato attraverso dei parametri caratteristici, conferendo a ciascun atomo un potenziale di interazione efficace che meglio lo rappresenta all'interno della struttura, infatti, a causa dei molteplici legami ed interazioni possibili chimicamente, non è escluso che atomi uguali siano caratterizzati da diverso potenziale.

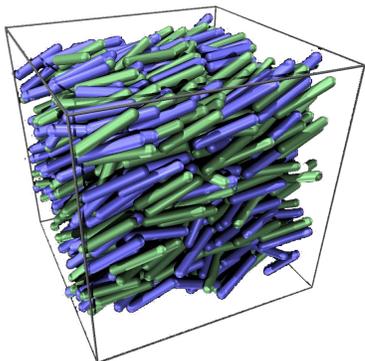


Figura 3.1: Transizione di fase nematica ottenuta mediante simulazione di cilindri rigidi.

Ad una prima analisi superficiale, l’approccio di dinamica molecolare “full-atom” sembrerebbe essere la migliore tecnica possibile per simulare il comportamento di un sistema atomico, poiché esso permetterebbe di conoscere molti più dettagli chimici rispetto ad una sua modellizzazione più semplicistica. Ma nella pratica, per colpa della presenza di un gran numero di gradi di libertà che debbono essere presi in considerazione (data la frequentissima necessità di considerare anche la presenza di acqua e ioni in soluzione), la dinamica molecolare risulta computazionalmente molto esigente [43] anche per la simulazione di piccoli sistemi di  $N_{sys} = 20k$  atoms come quelli trattati in questa sede, formati da doppi filamenti di DNA corti immersi in soluzione acquosa, limitando di fatto le scale di lunghezza e tempo che possono essere investigate. Un altro fattore discriminante

in tale analisi è che la grande quantità degli oggetti da simulare non riesce a riprodurre efficacemente la termodinamica del sistema stesso per colpa della reversibilità della dinamica molecolare, impedendo di fatto il raggiungimento in tempi finiti di una fase nematica come riscontrato nel lavoro di Maffeo et al. [21] e ne tantomeno di una struttura termodinamicamente stabile.

In un modello “coarse-grained” invece, la descrizione del DNA può essere semplificata riducendo in modo significativo il numero di gradi di libertà, considerando l’effettiva interazione tra unità “coarse-grained” al posto della più dettagliata interazione “full-atom”: la sfida e la difficoltà di tale modello è includere il minimo livello di dettaglio richiesto per catturare tutte le proprietà fisiche delle strutture da analizzare. I parametri dei “force-field”, che in questo caso non modellizzano le interazioni tra atomi bensì tra le unità che costituiscono il modello, utilizzati nell’approccio “coarse-grained” sono usualmente scelti in accordo con i dati sperimentali a disposizione con un approccio di tipo “bottom-up” [35] o viceversa mediante quello “top-down” [37].

### 3.1 Approccio “Bottom-Up”

In ambito informatico la strategia di programmazione “bottom-up”, in italiano traducibile dall’espressione “dal basso verso l’alto”, è una tecnica nella quale le singole parti che costituiscono un sistema sono esaminate nel dettaglio. Successivamente tutti questi costituenti elementari vengono connessi tra loro a formare delle parti più grandi e così via, fino ad arrivare alla descrizione completa dell’intero sistema.

Nell’approccio “bottom-up”, applicato al contesto delle simulazioni di dinamica molecolare, i parametri che definiscono i “force-field” sono settati dalla diretta comparazione con le

simulazioni atomistiche o con il confronto dei dati provenienti dalla struttura atomica: Proprio per questi motivi tale modello si presta molto bene alla riproduzione delle proprietà di conformazione strutturare di doppi filamenti di DNA.

Purtroppo l’approccio “bottom-up” solitamente fallisce nella riproduzione del comportamento del singolo filamento, nello studio del fenomeno dell’ibridizzazione ed in generale nel caratterizzare le proprietà termodinamiche del duplex, come ad esempio nel calcolo della sua temperatura di fusione. Tale tipo di schema concettuale è stato utilizzato nel lavoro di Maffeo [21] per stimare la forza di stacking tra due duplex di DNA interagenti attraverso le basi estremali partendo dal modello full-atom, anche se nello stesso lavoro si evidenzia l’impossibilità di raggiungimento di una fase nematica stabile per un sistema costituito da circa 500 doppi filamenti con la conseguente impossibilità dello studio delle sue proprietà termodinamiche.

### 3.2 Approccio “Top-Down”

Viceversa nella teoria dell’informazione, il modello di programmazione “dall’alto verso il basso” chiamato approccio “top-down”, prevede un’analisi del problema che parte dapprima con l’esaminare il sistema nel suo insieme, formulando una visione generale e successivamente suddivide il sistema in parti più piccole, raffinando via via la modellizzazione con l’aggiunta di maggiori dettagli.

Contrariamente a quanto avviene nel modello “bottom-up”, nell’approccio “top-down” i force-field sono disegnati per riprodurre con ragionevole accuratezza una larga serie di proprietà, come ad esempio la temperatura di fusione dell’elica, compatibilmente con quanto ottenuto in via sperimentale.

L’approccio “top-down” risulta perciò adeguato nello studio della termodinamica dei sistemi composti da doppi filamenti di DNA o da polimeri in generale, in particolare per caratterizzarne le transizioni di fase legate ai processi di “self-assembly”.

In generale, però, il range di applicazione di tale metodo è veramente molto vasto. In recenti articoli è stato proposto per studiare la renaturazione/denaturazione di acidi nucleici, ovvero quei processi che analizzano lo srotolamento del dDNA in due filamenti singoli di ssDNA e viceversa [39], nella formazione di hairpin mediante l’utilizzo della dinamica Browniana (processo che porta al ripiegamento su se stesso di un singolo filamento di acido nucleico) [40] e nella formazione della giunzione di Holliday (un assemblamento di singoli filamenti a formare una particolare struttura a quattro eliche) [37]. Inoltre come già accennato tale approccio è stato utilizzato per la caratterizzazione della termodinamica dei “duplex” [41], nello studio del fenomeno dell’“overstretching”, nel processo di gelificazione di nanoparticelle di DNA rivestito [42] e non ultimo nell’auto assemblamento delle doppie eliche corte [20].

Tra i vari approcci possibili per l’analisi termodinamica del sistema, è possibile utilizzare la modellizzazione “top-down” attraverso una schematizzazione coarse-grained dei singoli cos-

tituenti, nel quale ogni molecola di DNA è vista come una singola struttura rigida di forma cilindrica. Nell'articolo [19] viene proposto esattamente questo approccio, implementato mediante simulazioni Monte Carlo, ed ideato per esplorare le caratteristiche fisiche dell'equilibrio di polimerizzazione (“end-to-end stacking”) dei doppi filamenti corti di DNA immersi in soluzione acquosa. In questo modello i doppi filamenti sono assimilati a delle superquadriche con una forma quasi-cilindrica, dotate di due siti di interazione su ciascuna delle due basi, in modo da rappresentare le interazioni idrofobiche attrattive tra i “duplex”. Questo modello è stato studiato sia numericamente, mediante simulazione Monte Carlo negli ensemble a temperatura e pressione costante, e sia teoricamente utilizzando un appropriato funzionale di energia libera.

### 3.3 Il nostro approccio

Come abbiamo già accennato in precedenza, in questa tesi ci proponiamo di mostrare il ruolo giocato dal grado di piegatura, “bending”, nella formazione della fase liquido-cristallina in soluzioni ad alta concentrazione di DNA corti.

Dalla teoria sviluppata da Onsager-Bolhuis-Frenkel [14, 15] si evinceva come la formazione di fasi liquido-cristalline fosse dovuta sostanzialmente al solo contributo del volume escluso proprio dei doppi filamenti di DNA, essenzialmente assimilabili ad strutture cilindriche rettilinee molto allungate. A causa di ciò, il rapporto tra le dimensioni caratteristiche del cilindro, “aspect ratio” definito come  $X_0 = L/D$ , gioca un ruolo cruciale nella caratterizzazione del diagramma di fase della soluzione ed è stato possibile mostrare come al di sotto di un valore critico di  $X_0 = 4.7$  la transizione di fase, dovuta a soli contributi di volume, non fosse più permessa al sistema per nessuna densità. Contrariamente alle aspettative, il team di Nakata e Zanchetta [16] nel 2007, ha riscontrato la formazione di fasi nematiche anche in segmenti corti di DNA con aspect ratio compresi tra  $X_0 = 1.0 \div 3.3$  rendendo necessaria una nuova interpretazione del fenomeno. A queste grandezze, infatti, diventano rilevanti i contributi attrattivi idrofobici e di “base-stacking” tra le basi terminali del polimero aggiungendosi al contributo legato al volume escluso.

E' stata quindi necessaria una revisione computazionale al modello OBF, attuata da Glaser et al. [25], mediante l'introduzione di un potenziale attrattivo sulle basi del cilindro che approssima il polimero. Essi hanno mostrato computazionalmente come un sistema del genere potesse transire in diverse fasi stabili, passando da uno stato isotropico ad una fase nematica ad ordinamento maggiore e successivamente ad una struttura colonnare liquido-cristallina, sia nel caso di “short-duplex” DNA con  $X_0 \gtrsim 1.0$  che con strutture a disco caratterizzate da “aspect-ratio” particolarmente ridotti,  $X_0 \lesssim 1.0$ . L'introduzione del potenziale attrattivo sulle basi estremali è risultata una scelta consistente con gli esperimenti, e la sfida successiva è stata quella di stimare le forze responsabili del fenomeno. In questo contesto il lavoro di De Michele [19] viene introdotto come modifica del modello di Glaser mediante la sostituzione

degli oggetti a geometria cilindrica (HC) con superquadriche (SQ) di equazione:

$$f(x, y, z) = \left| \frac{x}{a} \right|^p + \left| \frac{y}{b} \right|^m + \left| \frac{z}{c} \right|^n - 1 = 0 \quad (3.1)$$

aventi  $m = n = 2$ ,  $p = 16$  e  $b = c$ . In questo modo l' "aspect-ratio" è ridefinito dalla geometria della superquadrica e risulta essere pari a  $X_0 = a/b$  compatibilmente con quanto già visto in precedenza. A dire il vero con la modellizzazione SQ dei doppi filamenti di DNA l'arrotondamento delle basi estremali dipende criticamente da  $X_0$  e quindi dalla scelta dei semiassi maggiori e minori della funzione  $f(x, y, z)$ . A causa della forma 3.1 con l'aumentare dell' "aspect ratio" le basi della superquadrica si stondono sempre di più influenzando sia nella determinazione del volume di ciascuna unità, sia nella capacità di aggregazione degli stessi. In realtà questo fattore non gioca un ruolo molto rilevante per le grandezze considerate e questo modello ha il pregio di poter utilizzare tali strutture per caratterizzare la dinamica di un sistema attraverso l'utilizzo di simulazioni "event-driven" grazie all'assenza di discontinuità nella geometria dei singoli costituenti [27].

In questo modo è stato possibile stimare per le energie libere responsabili del processo di "self-assembly" valori di circa  $\beta\Delta E_s = 5.56$  e  $\beta\Delta E_s = 6.67$ , del tutto in linea con i valori  $\beta\Delta E_s \approx 5 - 7$  previsti dall'articolo di Nakata [16]. Quindi a dispetto della semplificazione introdotta, è stato possibile mostrare la dipendenza della "volume fraction"  $\phi_N$  a cui avviene la transizione di fase dall'aspect ratio  $X_0$ .

A complicare ulteriormente il quadro della situazione, in un successivo lavoro di Nakata e Zanchetta [17] del 2010, è stata riscontrata una cruciale dipendenza della concentrazione critica a cui il sistema transisce rispetto alla sequenza di basi con cui sono costituiti gli acidi nucleici in soluzione acquosa raggiungendo variazioni fino al 70%, inoltre anche la chiralità globale della fase nematica del sistema è risultata strettamente correlata con la disposizione delle basi azotate.

La spiegazione di questo nuovo fenomeno non può essere addebitata alle sole forze idrofobiche di "stacking" che sono grossomodo indipendenti dai DNA considerati e ne tantomeno al volume escluso, perciò il modello a HC o SQ decorato con due siti attrattivi non risulta più adatto a caratterizzare la transizione di fase al variare della sequenza da cui sono costituiti gli acidi nucleici in soluzione.

Da studi effettuati mediante simulazioni full-atom sfruttando la dinamica molecolare, il team di Lankaš [54] in un lavoro del 2010, ha mostrato la presenza di un significativo grado di piegatura ("bending") delle sequenze in cui è presente il gruppo AATT. Poiché tale gruppo è presente in posizione centrale nelle sequenze palindrome a 12 basi azotate studiate in [16, 17], possiamo supporre che in questo caso il "bending" stesso giochi un ruolo cruciale nel determinare la concentrazione, e conseguentemente nella "volume-fraction"  $\phi_N$ , a cui avviene la transizione di fase nematica nel sistema.

	Sequence	Nick - name	N (bp)		$C_{IN}$ (mg/ml)	$ \rho $ range ( $\mu\text{m}$ )	H	$ \rho $ vs. $T$	$ \rho $ VS. $c$	sets
	...									
5	CGCGAATTCGCG	DD	12	BE	730	0.7 – 3	R	INC	-	1, 2
6	pCGCGAATTCGCG	pDD	12	BE	670	0.35 – 1	R	DEC	-	2
7	CGCGAATTCGCGp	DDp	12	BE	-	0.3 – 1	R	DEC	INC	2
8	GCGCTTAAGCGC	antiDD	12	BE	-	1 – 2	R	-	INC	-
9	GGAGTTTTGAGG + CCTCAAAACTCC	12mc	12	BE	770	0.7 – 2	R	-	-	-
10	ACCGAATTCGGT	ACC	12	BE	850	1 – 3	R	-	INC	2, 3
11	AACGAATTCGTT	AAC	12	BE	620	$\infty$	A	-	-	3
12	AATGAATTCATT	AAT	12	BE	500	1 – 2	L	-	-	3
13	AATAAATTTATT	allAT	12	BE	600	0.5 – 1	L	INC	-	3
14	CGGCGCGCCGG	allCG1	12	BE	650	1 – 3	L	-	-	3
15	CGCGCCGGCGCG	allCG2	12	BE	570	0.3 – $\infty$	L,R	INC,DEC	HI	-
16	GCGCGAATTCGC	sDD	12	SD	400	0.3 – 1	L	-	INC	2
	....									
28	long DNA	IDNA	150	-	160	2 – 4	L	INC	INC,DEC	-

Tabella 3.1: Estratto della tabella riportata nell’articolo PNAS di Nakata, Zanchetta et al. [17] in cui vengono considerate solamente sequenze lunghe  $N = 12$  basi azotate. Per convenzione adotteremo nel testo gli stessi nickname usati nell’articolo di riferimento, aggiungendo una convenzione cromatica (salvo diversamente indicato) per tre le sequenze principali che andremo ad analizzare nel seguito. Particolarmente importante è il valore critico della concentrazione  $C_{IN}$  a cui la soluzione isotropica transisce nello stato nematico che, come notiamo, dipende criticamente dalla sequenza considerata. Con  $H$  viene riportata l’elicità riscontrata nel sistema in fase nematica, mentre per le altre colonne si rimanda all’articolo [17].

L’idea alla base di questa tesi diventa quindi quella di caratterizzare il grado di piegatura di ciascuna struttura a doppia elica esaminata, restringendo l’attenzione in particolare alle sole sequenze palindrome di 12 basi azotate studiate nel lavoro di Nakata et al. [17] riportate in 3.1. Per fare questo sarà necessario investigare l’evoluzione temporale della struttura full-atom di ciascuna delle sequenze citate in precedenza per mezzo di simulazioni a dinamica molecolare. Traendo spunto dall’articolo di Dickerson [44], che definisce una grande quantità di metodi per analizzare le strutture delle doppie eliche di DNA mediante il software FREEHELIX, introdurremo delle osservabili create ad-hoc per il problema che ci consentiranno di caratterizzare il grado di piegatura di ciascun “duplex”. Ovviamente queste variabili verranno analizzate in funzione del tempo per controllare la stabilizzazione del sistema, inoltre, per ovviare a problemi legati ad effetti di taglia del sistema, ci spingeremo a simulare un sistema costituito da 18 sequenze che verranno analizzate globalmente e mediate per ottenere dei valori di riferimento del bending che utilizzeremo nel seguito.

Per rendere più comprensibile la lettura del testo decidiamo di introdurre uno schema

“red-green-blue” (RGB) per rappresentare le sequenze e gli oggetti “coarse grained”: in ogni contesto d’ora in poi l’oggetto più rettilineo (o supposto tale) verrà rappresentato con il colore rosso, mentre il più piegato con il blu.

Seguendo un approccio di tipo “bottom-up”, essendo partiti dalla dinamica dei singoli costituenti dell’acido nucleico per arrivare ad un grado di bending complessivo, andremo modellizzare la singola struttura di DNA mediante l’introduzione di un modello “coarse-grained” a doppio cilindro piegato (che chiameremo “hard bent cylinder”, HBC per brevità) con due siti attrattivi posti sulle basi estremali e caratterizzato dal grado di piegatura dipendente criticamente dall’osservabile introdotto in precedenza.

La seconda parte è perciò la parte decisiva, in cui utilizzeremo il modello a doppio cilindro piegato caratterizzato da due siti attrattivi i cui potenziali che ne mediano le interazioni verranno presi in prestito da i precedenti modelli ad HC e SQ [19, 20]. Nulla ci assicura che tali potenziali ben rappresentino le interazioni “self-assembly” di strutture HBC, ma è ragionevole pensare che in prima approssimazione forniscano le corrette dipendenze. Questo apre le porte ad eventuali ulteriori sviluppi del modello e rappresenta perciò il passo iniziale di un approccio “top-down” che consentirebbe un corretto settaggio dei “force-field”, in particolare dell’energia di “stacking”, per meglio rappresentare i dati ottenuti dagli esperimenti. In fine, per i problemi esposti in precedenza e dato che è necessaria un’analisi termodinamica del sistema, è molto conveniente in questo caso l’utilizzo di simulazioni Monte Carlo nell’ensemble NPT, che ci permettono di calcolare l’equazione di stato per ciascun sistema piegato.

L’analisi dei diagrammi di fase ottenuti per ciascuno dei sistemi HBC ci permetterà di osservare come un diverso grado di piegatura influisca fortemente nel determinare la concentrazione critica a cui il sistema transisce consentendoci di prevedere l’andamento della regione di coesistenza al variare del “bending” dei singoli doppi cilindri. In questo modo stiamo ancora una volta mettendo alla prova la potenza della “soft matter” nel controllo sempre più completo dei fenomeni fisici, infatti, grazie alla sola scelta della tipologia di sequenze di DNA di cui è costituito un sistema, potremmo deciderne a priori la densità critica di transizione fornendo anche un altro strumento utile per la creazione di nuovi materiali.

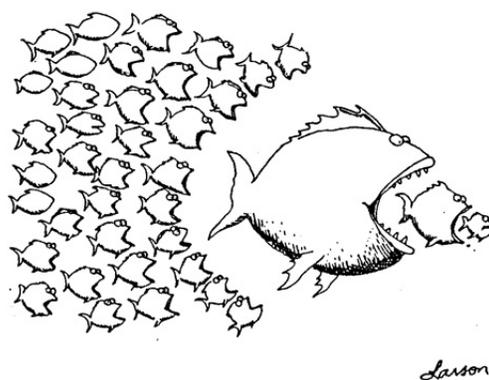


Figura 3.2: Schematizzazione del nostro approccio al problema: partendo dai singoli atomi arriveremo ad un modello coarse grained a doppio cilindro piegato che riprodurrà i dati ottenuti sperimentalmente solo dopo un successivo settaggio dei suoi stessi parametri.

## Parte II

# Dinamica Molecolare full-atom

## Capitolo 4

# Cenni di teoria sulla MD-Simulation

Abbiamo visto che la sempre maggiore potenza di calcolo dei recenti supercomputer ha reso possibile la simulazione della dinamica atomica di grandi sistemi, costituiti da almeno  $N > 20k$  atomi, in tempi relativamente brevi e permettendo agli scienziati di ottenere simulazioni di durata superiore ai  $10ns$  in circa 24 ore di simulazione [43]. Questo fatto rende tale tecnica sempre più utilizzata per l'analisi delle proprietà strutturali di grandi polimeri, perché a differenza degli esperimenti reali, possiamo controllare la dinamica di ciascun atomo (o di gruppi di essi) attraverso la definizione di opportune osservabili. La potenza del metodo risiede tutta nella definizione del “force-field” del sistema che negli anni è stato affinato per rispondere alle proprietà comportamentali delle molecole ottenute negli esperimenti.

La simulazione del comportamento di un sistema mediante la dinamica molecolare [5, 6], “Molecular-Dynamic Simulation” (MD), è una tecnica computazionale per determinare le proprietà di trasporto e di equilibrio di un sistema a molti corpi, le cui equazioni del moto non possono essere risolte analiticamente per via della grande quantità di interazioni e di condizioni iniziali a cui il sistema deve sottostare. In generale, quando ci si riferisce alla MD, si fa implicitamente riferimento all'integrazione numerica di equazioni classiche del moto: solo quando consideriamo moti traslazionali o rotazionali di atomi o molecole leggere ( $He$ ,  $H_2$ ,  $D_2$ ) oppure per analizzare vibrazioni con frequenza  $\nu$  tale che  $h\nu > k_B T$  ci dovremo preoccupare di una trattazione quantistica del problema. Nel nostro caso perciò, dato che andremo a simulare sistemi polimerici in soluzione acquosa, la trattazione classica del moto sarà più che sufficiente a descrivere accuratamente il problema.

Dal punto di vista scientifico, la simulazione MD è in molti aspetti simile ad un vero e proprio esperimento a laboratorio, perciò gli step da seguire per ottenere una simulazione realistica sono correlati alla sua controparte reale:

1. Per prima cosa occorre preparare il campione da simulare, costituito da  $N$  particelle, risolvendo le equazioni di Newton affinché le proprietà temporali siano quelle di equilibrio, in questo caso si parla di *processo di equilibratura del sistema*;

2. Si simula l'evoluzione del sistema monitorando le grandezze termodinamiche importanti: Energia, Temperatura, Volume, Pressione e Numero di particelle devono essere costanti a seconda dell'ensemble statistico che si vuole esplorare. Questa è la vera parte di *simulazione di dinamica molecolare* in cui avviene la produzione dei dati da analizzare;
3. A simulazione conclusa, dopo aver definito delle opportune osservabili, si misurano le grandezze fisiche di interesse al variare del tempo. Questa quindi è la parte di *analisi dati* che ci permetterà di trarre delle conclusioni alla fine del nostro esperimento computazionale.

Esistono diverse tecniche di integrazione delle equazioni del moto, in particolare viene maggiormente utilizzato l'*algoritmo di Verlet* per via della sua semplicità e stabilità rispetto alle conservazioni di energia del sistema [5]. Tutti gli algoritmi symplettici che preservano la forma dello spazio delle fasi hamiltoniano partono necessariamente da un'espansione in serie di Taylor della posizione  $\mathbf{r}$  al tempo  $t$ , considerando istanti successivi e precedenti ed approssimando al terzo ordine abbiamo che:

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t) \Delta t + \frac{1}{2m} \mathbf{F}(t) \Delta t^2 + \frac{1}{3!} \mathbf{r}''' \Delta t^3 + \mathcal{O}(\Delta t^4) \quad (4.1)$$

$$\mathbf{r}(t - \Delta t) = \mathbf{r}(t) - \mathbf{v}(t) \Delta t + \frac{1}{2m} \mathbf{F}(t) \Delta t^2 - \frac{1}{3!} \mathbf{r}''' \Delta t^3 + \mathcal{O}(\Delta t^4) \quad (4.2)$$

che sommate ci permettono di ottenere per l'istante successivo  $t + \Delta t$ :

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \frac{1}{m} \mathbf{F}(t) \Delta t^2 + \mathcal{O}(\Delta t^4) \quad (4.3)$$

rappresenta un'espressione per determinare le nuove posizioni conoscendo le vecchie e la forza che agisce sull'atomo. Tale espressione contiene un errore del quarto ordine nel tempo e risulta quindi è molto precisa ed efficiente, mentre per quanto concerne le velocità sottraendo le espressioni 4.1 e 4.2 otteniamo:

$$\mathbf{v}(t) = \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t - \Delta t)}{2\Delta t} + \mathcal{O}(\Delta t^4) \quad (4.4)$$

con un errore del secondo ordine. Questo algoritmo, ripetuto iterativamente a step temporali successivi per ciascuno degli atomi presenti, porta alla computazione della dinamica molecolare dell'intero sistema considerato. Ovviamente sono presenti altre tecniche di integrazione per spostare l'errore ad ordini successivi, come ad esempio l'algoritmo di predizione-correzione, anche se in questa sede siamo interessati esclusivamente all'algoritmo sopra esposto.

Poiché l'algoritmo prevede la risoluzione di equazioni differenziali del secondo ordine è necessario conoscere i dati iniziali da cui far partire l'evoluzione per poter stimare correttamente la traiettoria del moto. Se il sistema è costituito da  $N$  particelle sarà necessario avere

a disposizione  $6N$  condizioni iniziali,  $3N$  delle quali determinate dalla posizione e salvate nel file \*.pdb descritto nel capitolo precedente, e le restanti  $3N$  coordinate delle velocità. Ovviamente non è possibile stabilire con esattezza le velocità delle singole particelle, perciò esse verranno in generale assegnate casualmente a ciascuno degli atomi secondo una distribuzione di Maxwell-Boltzman ad una data temperatura  $T$ :

$$p(v_i) = \sqrt{\frac{m_i}{2\pi k_B T}} e^{-\frac{m_i v_i^2}{2k_B T}} \quad (4.5)$$

dove  $k_B$  è la costante di Boltzmann,  $m_i$  e  $v_i$  massa e velocità della particella  $i$ -esima e  $T$  la temperatura a cui si trova il sistema.

In generale le osservabili che si vogliono misurare vengono calcolate come espressamente dipendenti dalla posizione e dalla quantità di moto dei singoli costituenti del sistema, per esempio nel caso della temperatura può essere utile introdurre il principio dell'equipartizione dell'energia che ci consente di calcolare esplicitamente la grandezza in funzione delle velocità atomiche:

$$\left\langle \frac{1}{2} m v_\alpha^2 \right\rangle = \frac{1}{2} k_B T \quad \rightarrow \quad T(t) = \sum_{i=1}^N \frac{m_i v_i^2}{k_B N_f} \quad (4.6)$$

in questo caso la fluttuazione della grandezza sarà dell'ordine di  $N_f^{-1/2}$  e perciò per ottenere delle stime attendibili per la temperatura, dovremo valutare tale grandezza su tempi scelti in modo tale da mediare sul maggior numero di fluttuazioni possibili.

Molti sono i programmi che si occupano di simulare la dinamica molecolare anche se in questa sede opteremo per l'utilizzo del software freeware GROMACS, risultato essere molto performante [43] nella simulazione di sistemi dell'ordine di  $\sim 10 \div 100k$  atomi rispetto ai suoi principali concorrenti: AMBER, CHARMM, LAMMPS e NAMD.

## 4.1 Software GROMACS

Come già accennato, in questa sede andremo a sfruttare il motore di calcolo di dinamica molecolare GROMACS, acronimo di "Groningen Machine for Chemical Simulations". Al momento della scrittura è stata da poco rilasciata la versione 4.6.1 del software, anche se la vera novità è stata introdotta nella 4.6 con l'implementazione delle librerie CUDA per poter eseguire simulazioni di dinamica molecolare sfruttando la potenza di calcolo delle GPU nVidia di recente costruzione. Il supporto alle librerie MPI per la parallelizzazione su normali CPU è invece presente fin dalle prime versioni, anche se dalla versione 4.5, grazie all'introduzione dei "threads", non è più necessario invocarlo esplicitamente per eseguire dei run su singoli nodi o

Quantity	Symbol	Unit
length	$\mathbf{r}$	$nm = 10^{-9}m$
mass	$m$	$u = 1.6605402 \cdot 10^{-27}kg$
time	$t$	$ps = 10^{-12}s$
charge	$q$	$e = 1.60217733 \cdot 10^{-19}C$
temperature	$T$	$K$

Derived Quantity	Symbol	Unit
energy	$E, V$	$kJ mol^{-1}$
force	$\mathbf{F}$	$kJ mol^{-1} nm^{-1}$
pressure	$p$	$kJ mol^{-1} nm^{-3} = 1.66054 \cdot 10^6 Pa$
velocity	$v$	$nm ps^{-1} = 1000m s^{-1}$
dipole moment	$\mu$	$e nm$
electric potential	$\Phi$	$kJ mol^{-1}$
electric field	$E$	$kJ mol^{-1}$

Tabella 4.1: Tabelle delle grandezze di base e derivate utilizzate dai programmi del pacchetto GROMACS. Anche in questa tesi utilizzeremo le stesse convenzioni.

su processori multicore<sup>1</sup>.

GROMACS implementa delle tecniche di risoluzione delle equazioni di Newton classiche ed un gran numero di strumenti per l'analisi delle traiettorie risultanti, costituendo un pacchetto completo per lo studio di simulazioni MD.

La risoluzione delle equazioni differenziali del moto sono affidate ad un algoritmo di integrazione standard denominato "leap-frog" poiché prevede il calcolo delle posizioni e velocità a tempi alterni:

$$\mathbf{v} \left( t + \frac{1}{2} \Delta t \right) = \mathbf{v} \left( t - \frac{1}{2} \Delta t \right) + \frac{\Delta t}{m} \mathbf{F}(t) \quad (4.7)$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \Delta t \mathbf{v} \left( t + \frac{1}{2} \Delta t \right) \quad (4.8)$$

che produce traiettorie equivalenti a quelle prodotte dall'algoritmo di Verlet descritto in precedenza poiché la relazione che lega la posizione a tempi successivi può essere scritta, sostituendo la 4.7 nella 4.8, come:

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \frac{1}{m} \mathbf{F}(t) \Delta t^2 + \mathcal{O}(\Delta t^4) \quad (4.9)$$

del tutto identica alla 4.3.

Oltre a questo algoritmo, caratterizzato da una precisione del terzo ordine per il calcolo delle posizioni ed inoltre reversibile temporalmente, GROMACS offre la possibilità di sfruttare

<sup>1</sup>Anche se in realtà a noi è risultato più performante esplicitare il numero di core con cui parallelizzare anche nel caso di CPU multicore.

un altro algoritmo di integrazione simplettico denominato “velocity-Verlet” che, seppur stia diventando il nuovo standard nella dinamica molecolare, scegliamo di non utilizzare poiché non pienamente supportato da GROMACS e poiché il “leap-frog” risulta abbastanza accurato per gran parte delle simulazioni MD.

Per quanto riguarda i “Force-Field” usati essi sono tutti conservativi e dipendenti dalla posizione dei soli atomi, il che implica che gli elettroni negli atomi vengono considerati tutti allo stato fondamentale e la loro dinamica cambia istantaneamente quando cambia la posizione atomica (approssimazione di Born-Oppenheimer). Inoltre i F-F sono additivi ed approssimati e possono essere modificati dall’utente per aggiustare i parametri di alcuni atomi, per sondare ad esempio altri comportamenti oppure far combaciare i risultati sperimentali con le simulazioni. Un’altra caratteristica da tenere presente è che le interazioni a lungo raggio sono escluse dal calcolo mediante alcune tecniche di “cut-off” ed anche le condizioni al bordo non sono realistiche, questo perché un sistema costituito da  $10^k$  atomi avrebbe un gran numero di interazioni con la superficie del box in cui è contenuto e per fare delle simulazioni realistiche del “bulk” occorre introdurre necessariamente delle approssimazioni. Questo porta ovviamente a dei problemi legati alla dimensione del box in cui avviene la simulazione ed occorrerà analizzare di volta in volta i risultati ottenuti per non cadere nella misurazione di grandezze errate (per esempio attraverso le condizioni periodiche al bordo la molecola potrebbe vedere se stessa, condizione non riscontrabile in natura).

Per quanto riguarda il calcolo e l’analisi dati useremo la convenzione adoperata nel manuale del software, in cui le unità di base sono  $nm$ ,  $ps$ ,  $K$ , carica elettrica ( $e$ ) ed unità di massa atomica a cui si aggiunge il fattore di conversione tra le quantità meccaniche ed elettriche:

$$V = f \frac{q^2}{r} \quad \text{oppure} \quad F = f \frac{q^2}{r^2} \quad (4.10)$$

pari a  $f = \frac{1}{4\pi\epsilon_0} = 138.935485 kJ mol^{-1} nm e^{-2}$  mentre tutte le altre unità sono riportate in tabella 4.1.

Pur supportando numerose tipologie di box all’interno del quale fare evolvere il sistema, noi considereremo esclusivamente una geometria cubica di lato  $a$ , con cella unitaria definita dai vettori di base  $\mathbf{a}_1 = (a, 0, 0)$ ,  $\mathbf{a}_2 = (0, a, 0)$  ed  $\mathbf{a}_3 = (0, 0, a)$  ritenuta la migliore per i scopi della nostra analisi, applicando per una generica funzione  $f(\mathbf{r})$  e per la posizione stessa degli atomi le condizioni periodiche al bordo di Born-von Karman in tre dimensioni  $f(\mathbf{r} + N\mathbf{a}_j) = f(\mathbf{r})$ , supponendo cioè che se una particella esce da un lato del box rientra dal lato opposto. In questo modo il sistema viene replicato infinite volte attorno

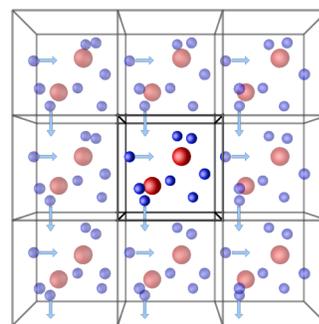


Figura 4.1: Periodic Boundary Conditions” di Born von Karman, se un atomo esce da un lato della scatola rientra da quello opposto.

al box principale, come mostrato in figura, creando delle immagini che circondano il sistema originale.

Avendo una struttura del genere è raccomandabile scegliere come distanza massima per il cut-off delle interazioni la metà del lato più piccolo del box in cui vogliamo simulare il sistema, ovvero dovrà essere valida la disuguaglianza  $R_{cut-off} < \frac{1}{2}a$ . Questo fatto è di grande importanza per far sì che la molecola non interagisca con se stessa attraverso l'applicazione delle condizioni periodiche al bordo. Per quanto concerne la discussione di altre possibili tipologie di box e cut-off rimandiamo al manuale online del software.

## 4.2 Trattazione delle forze e delle interazioni tra particelle

Il pacchetto software di dinamica molecolare GROMACS ci mette a disposizione un gran numero di modelli per trattare le interazioni tra le particelle del sistema che andremo a considerare. In particolare le forze che regolano l'interazione tra gli atomi dipendono criticamente dalla scelta del “force-field” che andremo ad utilizzare per integrare le equazioni del moto, mentre la distanza tipica a cui le interazioni verranno troncate dipende dalla scelta dello schema di “cut-off” che scegliamo per la nostra simulazione.

Con il termine “force-field” (F-F) ci riferiamo all'intero set di parametri e di forme funzionali utilizzate per descrivere *l'energia potenziale di un sistema di particelle*. I parametri contenuti nei F-F sono calcolati per meglio rappresentare determinate proprietà fisiche cercando di far combaciare i risultati sperimentali con le simulazioni. Questo implica un continuo affinamento e una sempre maggiore vastità di parametri da considerare nelle simulazioni, rendendo di volta in volta necessario aggiornare i parametri per far fronte a nuove scoperte scientifiche.

Grazie ad un continuo feedback [32, 46] con i dati sperimentali, i recenti FF descrivono egregiamente la maggior parte delle caratteristiche strutturali di grandi biomolecole, mentre invece falliscono quando si tratta di analizzare la termodinamica di sistemi molto complessi [21].

Nella dinamica molecolare “full-atom” il force field contiene tutti i parametri per descrivere ciascuno degli atomi presenti nella struttura, mentre per la simulazione di oggetti “coarse-grained” vengono usate rappresentazioni più astratte per poter investigare a tempi lunghi il comportamento del sistema. In ciascuno dei due casi il “force field” definisce un campo di forze additivo la cui decomposizione energetica può in generale essere scritta sempre separando due contributi:

$$V_{total} = V_{bonded} + V_{non-bonded} \quad (4.11)$$

in cui nel primo termine si fa riferimento a contributi energetici determinati da legami covalenti

interatomici

$$V_{bonded} = V_{bonds} + V_{angle} + V_{dihedral} \quad (4.12)$$

mentre nel secondo termine vengono raccolti tutti i contributi di legami non covalenti, quali le interazioni elettrostatiche e le forze di Van Der Waals:

$$V_{non-bonded} = V_{electronic} + V_{VanDerWaals}. \quad (4.13)$$

Computazionalmente è molto più dispendioso il calcolo delle energie dei legami non covalenti, poiché ciascun atomo possiede molte più interazioni possibili, mentre per i legami covalenti, molto più forti, i contributi di legame si restringono solo a gruppi di atomi primi vicini.

Numerose tipologie di “force-field” sono state sviluppate nel tempo, citiamo fra tutte a titolo di esempio CHARMM, GROMOS, OPLS ed AMBER, che in questa sede decidiamo di utilizzare nella sua versione più recente, denominata AMBER03 [32] un’evoluzione dell’AMBER-parm99 utilizzato nelle simulazioni di Maffeo [21].

#### 4.2.1 fAMBER03

I force-field AMBER, acronimo di “Assisted Model Building with Energy Refinement”, sono una famiglia di F-F utilizzati in ambito di dinamica molecolare “full-atom” di grandi molecole organiche sviluppato originariamente dal team di Peter Kollman dell’Università della California a San Francisco. Al momento della scrittura è disponibile nella release open-source AMBER03[32], un’evoluzione dell’AMBER-parm99 utilizzato nelle simulazioni del team di Maffeo [21] che ben si presta all’analisi della struttura di proteine ed acidi nucleici. Recentemente è stata messa a disposizione una nuova versione a pagamento denominata AMBER10.

Come per tutti i FF, anche in AMBER il contributo energetico è separabile in diversi contributi dovuti alle interazioni covalenti ( $V_{bonded}$ ) ed a quelle non covalenti ( $V_{non-bonded}$ ) tra atomi. Per le prime i contributi della 4.12 si scrivono esplicitamente come:

$$V_{bonds} = \sum_{bonds} K_r (r - r_{eq})^2 \quad (4.14)$$

$$V_{angles} = \sum_{angles} K_\theta (\theta - \theta_{eq})^2 \quad (4.15)$$

$$V_{dih} = \sum_{dihedrals} \frac{1}{2} V_n [1 + \cos(n\phi - \gamma)]. \quad (4.16)$$

Col primo termine,  $V_{bonds}$ , viene espressa l’energia di legame tra atomi legati tra loro mediante legami covalenti rappresentata mediante un potenziale armonico. Questa modellizzazione

risulta una buona approssimazione vicino alla distanza di equilibrio del legame, ma sottostima le interazioni al crescere della separazione atomica. Il secondo termine rappresenta l'energia di legame dovuta alla geometria degli orbitali atomici coinvolti nei legami covalenti instaurati dall'atomo mentre il terzo,  $V_{dihedrals}$ , fornisce una misura dell'energia torsionale relativa agli angoli diedri espressa in serie di Fourier per far fronte alla possibilità che un singolo legame possa avere più termini di questo tipo.

Infine il contributo relativo ai legami non covalenti 4.13 può essere separato in due termini:

$$V_{VanDerWaals} = \sum_{j=1}^{N-1} \sum_{i=j+1}^N \epsilon_{i,j} \left[ \left( \frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{r_{0ij}}{r_{ij}} \right)^6 \right] \quad (4.17)$$

$$V_{electronic} = \sum_{j=1}^{N-1} \sum_{i=j+1}^N \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}. \quad (4.18)$$

Il termine  $V_{VanDerWaals}$  è proprio il contributo delle forze di Van Der Waals e viene espresso mediante un potenziale di Lennard-Jones i cui parametri sono calcolati stimando la distanza di equilibrio  $r_{0ij}$  e la profondità della buca di interazione  $\epsilon_{i,j}$ , mentre la presenza del fattore 2 assicura che la distanza di equilibrio è proprio  $r_{0ij}$ . Infine  $V_{electronic}$  è il contributo delle interazioni elettrostatiche tra i vari atomi. Tutte le costanti introdotte sono definite opportunamente nei database AMBER, come indicato in [46].

### 4.2.2 Modello TIP3P

Anche per simulare il comportamento dell'acqua sono necessari dei modelli che descrivono accuratamente le interazioni della molecola con le altre molecole del solvente ed anche con il soluto stesso. Spesso infatti le interazioni di molecole non polari sono a carattere idrofobico e la dinamica dell'acqua influenza grandemente il comportamento del sistema rendendo fondamentale la corretta scelta dei parametri che modellizzano la molecola.

Per quanto concerne la simulazione della dinamica delle molecole d'acqua decidiamo quindi di sfruttare il modello TIP3P come fatto anche nel lavoro di Maffeo [21]. Questo modello caratterizzato dalla presenza di tre siti di interazione e risulta molto accurato per i nostri fini di analisi.

Il modello TIP3P descrive l'interazione elettrostatica della molecola d'acqua mediante un potenziale elettrostatico di tipo Lennard-Jones [45, 47], utilizzato in un modello a tre siti di interazione, ciascuno per ogni posizione atomica:

$$V_{ab}^{TIP3P} = \sum_i^{on\ a} \sum_j^{on\ b} \frac{k_c q_i q_j}{r_{ij}} + \frac{A}{r_{OO}^{12}} - \frac{B}{r_{OO}^6} \quad (4.19)$$

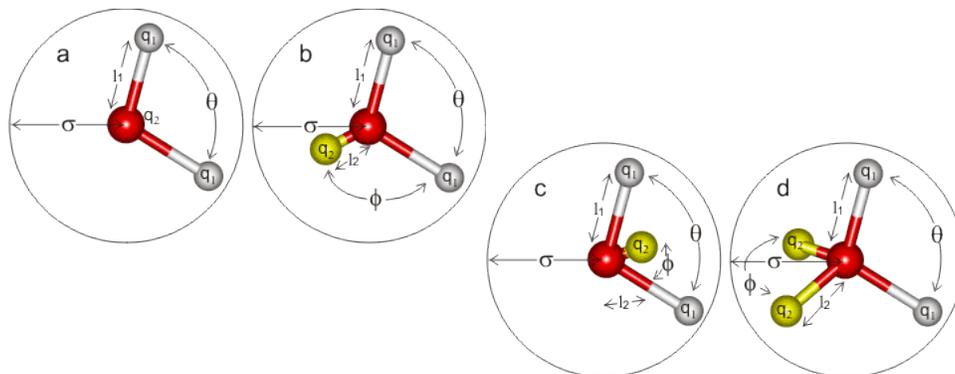


Figura 4.2: Differenti tipologie di modelli a più siti di interazione per l' $H_2O$ . Il modello preso in esame in questo testo è il primo a destra, a due siti. Gli altri sono riportati a titolo esemplificativo e corrispondono ai modelli a 4, 5 e 6 siti di interazione.

dove  $k_C = \frac{1}{4\pi\epsilon_0} = 332.1 \text{Åkcal/mol}$  è la costante elettrostatica,  $q_i$  sono le cariche parziali relative alla carica degli elettroni, che nel modello TIP3P vengono settate pari a:

$$q_O = -0.834e \quad q_H = +0.417e \quad (4.20)$$

mentre  $r_{ij}$  è la distanza tra due siti atomici carichi, ovvero:

$$r_{OH} = 0.9572\text{Å} \quad r_{OO} = 1.5139\text{Å}. \quad (4.21)$$

Attraverso la scelta di questi parametri l'angolo al vertice risulta pari a  $\theta_{HOH} = 104.52^\circ$  ovvero quello osservato dai dati sperimentali, a differenza della forma tetraedrica perfetta del modello SPC (Simple Point Charge) in cui però, come nel TIP3P, non vengono parametrizzati gli effetti dovuti alla presenza dei "lone-pair" in due dei vertici del tetramero. I parametri  $A$  e  $B$  sono i parametri di L-J, che in questo modello vengono settati a:

$$A = 582.0 \cdot 10^{-3} \text{kcalÅ}^{12}/\text{mol} \quad B = 582.0 \text{kcalÅ}^6/\text{mol} \quad (4.22)$$

Utilizzando modelli più complessi a 4 o più siti di interazione (TIP4P, TIP4P-Ew, TIP5P, etc.) potremmo ottenere simulazioni più realistiche ma computazionalmente molto più esigenti, poiché con tali modelli vengono aggiunti dei siti interagenti che schematizzano la presenza dei lone-pair e delle conseguenti ridistribuzioni di carica. In questa sede desideriamo investigare solo le proprietà strutturali dei frammenti di DNA, non siamo interessati alle dinamiche di interazione con il solvente, perciò il modello TIP3P viene scelto per la sua comprovata efficienza computazionale e per la sua semplicità mantenendosi in linea con quanto fatto da Maffeo [21].

### 4.2.3 Trattazione del “cut-off” delle interazioni

Le interazioni tra particelle vengono calcolate attraverso la scelta di particolari tecniche di “cut-off”. Poiché il box è finito occorre fare in modo che la distanza a cui tronchiamo le interazioni sia inferiore alla metà della dimensione minima del box che contiene la molecola da esaminare, per fare in modo che essa non interagisca con se stessa. Il metodo standard mediante il quale GROMACS si occupa del “cut-off” è denominato “group scheme”, che prevede la creazione di gruppi di atomi che interagiscono se il baricentro degli stessi cade all’interno del raggio di “cut-off”. Questo metodo è particolarmente efficiente quando il sistema, come nel nostro caso, è costituito in gran parte da molecole d’acqua. A partire dalla versione 4.6, contestualmente all’aggiunta del supporto delle librerie CUDA, è stato introdotto un altro metodo per la trattazione del “cut-off” denominato “Verlet-scheme”, differente rispetto allo schema precedente principalmente per il fatto che i gruppi non sono statici ma vengono periodicamente aggiornati. Questo risulta computazionalmente efficiente quando vengono utilizzate le GPU per l’esecuzione del run ed è anche l’unico schema di trattazione dei “cut-off” disponibile per questo tipo di simulazione.

Per quanto riguarda invece la trattazione del troncamento dell’interazione elettrostatica, decidiamo di utilizzare il metodo migliore a nostra disposizione, il metodo della somma di Ewald, che in uno spazio suddiviso mediante una griglia può essere implementato con la tecnica PME (Particle Mesh Ewald). A discapito di una minore velocità di esecuzione ci consente di ottenere risultati migliori rispetto a un cut-off standard il quale tronca brutalmente l’interazione ad una data distanza.

Tale metodo, a differenza del “cut-off” standard nello spazio diretto che impone il potenziale di interazione nullo al di fuori di un certo range, sfrutta il fatto che il solvente in cui è immerso il nostro segmento di DNA schermi parzialmente l’interazione a lungo raggio con un certo potenziale medio, introdotto mediante l’inserimento artificiale di cariche fittizie all’interno del sistema.

La particolarità del metodo PME è che esso prevede la separazione del contributo del potenziale in tre parti:

$$V_{electronic}^{PME} = V_r + V_m + V_0 \quad (4.23)$$

ovvero come somma di una parte reale nello spazio diretto  $V_r$ , una parte reciproca nello spazio immaginario (effettuata in trasformata di Fourier)  $V_m$ , ed un termine costante, chiamato self-term, un termine di correzione che annulla l’auto-interazione delle contro cariche introdotte artificialmente dal metodo:

$$V_r = \frac{1}{2} \sum_{i,j}^{N'} \sum_{\mathbf{n}} \frac{k_c q_i q_j}{r_{ij,\mathbf{n}}} \operatorname{erfc}(\alpha r_{ij,\mathbf{n}}) \quad (4.24)$$

$$V_m = \frac{1}{2\pi V} \sum_{i,j}^N q_i q_j \sum_{\mathbf{m} \neq 0} \frac{e^{-\left(\frac{\pi|\mathbf{m}|}{\alpha}\right)^2 + 2\pi i \mathbf{m} \cdot (\mathbf{r}_i - \mathbf{r}_j)}}{|\mathbf{m}|^2} \quad (4.25)$$

$$V_0 = -\frac{\alpha}{\sqrt{\pi}} \sum_{i=1}^N q_i^2 \quad (4.26)$$

con  $V$  volume del box,  $\mathbf{n}$  vettore che identifica le celle del reticolo diretto,  $\mathbf{m}$  vettore del reticolo reciproco, la “error-function”  $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du$ .

Ulteriori dettagli del metodo sono trattati in [48, 49, 50] ed anche nel testo di Kittel [51]. Il metodo per trattare il troncamento del potenziale di Van Der Waals invece è un normale troncamento a “cut-off”, opzione standard in una simulazione con GROMACS ed entrambe le scelte del trattamento delle interazioni sono state fatte in linea con le scelte nell’articolo di Maffeo [21].

### 4.3 Set-up dei parametri della simulazione

Tutti i parametri con i quali andremo a simulare il nostro esperimento di “molecular dynamics” vengono salvati in un file di testo con estensione \*.mdp grazie al quale il software **grompp** del pacchetto GROMACS prepara il file \*.tpr con il quale lanciare la simulazione.

I parametri termodinamici con cui gestiamo il nostro sistema ovviamente saranno settati in modo da essere compatibili con quelli tipici di un esperimento in laboratorio, ovvero temperatura ambiente di  $T = 300K$  e pressione impostata a  $P = 1bar$ . Tali grandezze vengono periodicamente controllate durante la simulazione mediante la scelta di criteri di “coupling” che consentono di effettuare le simulazioni in determinati insiemi. Poiché fissiamo pressione, temperatura e numero di particelle, la nostra simulazione sarà effettuata nell’ensemble NPT e tali grandezze dovranno rimanere costanti. A dire il vero solo il numero di particelle contenute nel sistema non varia, poiché GROMACS mette a disposizione dei termostati per far fluttuare in modo controllato anche i valori di pressione e temperatura. Abbiamo scelto di usare questo tipo di approccio per riprodurre in modo più accurato possibile la situazione che si verifica sperimentalmente: tenendo presente che la concentrazione di DNA in soluzione viene fatta aumentare facendo evaporare acqua dal sistema [16] è ragionevole supporre che ci siano fluttuazioni in temperatura e pressione, oltre che in volume, durante la transizione di fase. In questo contesto le nostre simulazioni rappresenteranno una singola parte dell’esperimento, riproducendo solamente una situazione dinamica a determinate concentrazioni di soluto. Purtroppo pensare di effettuare uno studio in concentrazione del DNA variando il numero di acidi nucleici all’interno di un box a volume fissato, sebbene possa essere davvero molto interessante, è al di fuori della portata delle risorse hardware a nostra disposizione.

Per quanto riguarda la pressione decidiamo di utilizzare l'algoritmo chiamato "Berendsen pressure coupling", che riscalda ad periodicamente le coordinate atomiche e le dimensioni del box mediante una matrice  $\mu$  di forma:

$$\mu_{ij} = \delta_{ij} - \frac{n_{PC}\Delta t}{3\tau_P}\beta_{ij}\{P_{0ij} - P_{ij}(t)\} \quad (4.27)$$

dove  $n_{PC}$  è il numero di step dopo il quale la pressione viene riscalata e  $\beta$  compressibilità isoterma del sistema. Questo ci permette di controllare le fluttuazioni della pressione nel tempo definite come:

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau_P} \quad (4.28)$$

infatti la soluzione dell'equazione differenziale 4.28 implica un decadimento esponenziale della deviazione della pressione rispetto a  $P_0$  con costante di tempo  $\tau_P$ .

La temperatura invece, decidiamo di controllarla mediante un algoritmo analogo a quello di Berendsen a cui viene introdotta una correzione che assicura la conservazione dell'energia cinetica. Tale metodo viene chiamato "velocity-rescaling thermostat" e che viene implementato riscaldando le velocità (invece delle posizioni del caso del coupling della pressione) mediante la:

$$dK = (K_0 - K) \frac{dt}{\tau_T} + 2\sqrt{\frac{KK_0}{N_f}} \frac{dW}{\sqrt{\tau_T}} \quad (4.29)$$

dove  $K$  è l'energia cinetica,  $\tau_T$  la costante di tempo sul quale vengono riscalate le temperature,  $N_f$  il numero dei gradi di libertà e  $dW$  il processo di Wiener. Questo algoritmo ha gli stessi vantaggi del Berendsen, ovvero un decadimento delle temperature al primo ordine senza introdurre oscillazioni.

Fondamentale per ottenere delle simulazioni realistiche è la scelta dell'intervallo temporale con cui integrare le equazioni del moto. A seconda delle proprietà che vogliamo studiare dovremo settare il passo temporale  $\Delta t \rightarrow dt$  in modo tale da essere inferiore rispetto al tempo di rilassamento tipico delle vibrazioni del sistema che siamo interessati a studiare. Nel nostro caso il passo di integrazione temporale viene impostato a  $dt = 0.002ps = 2fs$  per un corretto funzionamento del termostato di Berendsen e perché risulta una buona scelta per determinare gli angoli di legame tra piani molecolari. Infatti tempo di rilassamento tipico di vibrazioni a media frequenza, come quelle legate a deformazioni angolari, torsione di angoli diedri, legami di Van Der Waals e Coulombiani a corto raggio, è pari a circa  $40fs$  e perciò la scelta dell'intervallo di integrazione  $dt = 2fs$  risulta adeguata.

In seguito vedremo che le simulazioni dovranno avere durata totale maggiore di  $t_{simulation} > 100ns$ , perciò gli step di integrazione verranno sempre settati in modo tale da essere  $\#step > 5 \cdot 10^7$ . In particolare le lunghezze temporali delle simulazioni che analizzeremo saranno da  $200ns$ ,  $300ns$  oppure  $400ns$  a seconda della tipologia di analisi che andremo a fare.

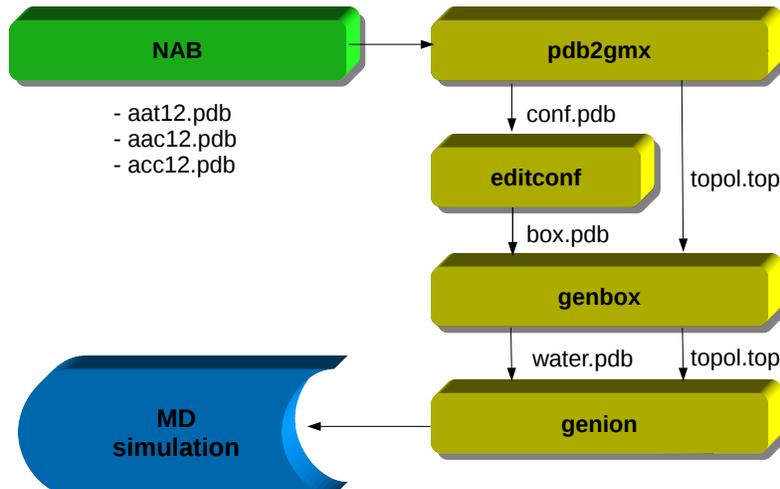


Figura 4.3: "Flow-chart" che mostra la sequenza di operazioni da effettuare per preparare il sistema alla simulazione a dinamica molecolare.

#### 4.4 Preparazione del sistema

Per la preparazione del sistema ci atterremo sostanzialmente allo schema mostrato in figura 4.3, che illustra la sequenza dei programmi da eseguire per completare il set-up dell'esperimento computazionale.

Poiché ci occuperemo della simulazione di segmenti di DNA palindromi, costituiti da 12 coppie di basi azotate come studiato in [16], avremo bisogno in primo luogo di costruire la struttura geometrica dell'acido nucleico. Per fare questo ci serviremo del software NAB (acronimo di "Nucleic Acid Builder") incluso nel pacchetto software AmberTools12, che consente di generare sequenze di B-DNA di qualsiasi tipo prendendo in input una stringa di caratteri che rappresentano la sequenza di basi azotate dell'acido nucleico e salvando la geometria creata su un file \*.pdb.

A questo punto verrà creata la topologia che conterrà le informazioni sui force-field utilizzati per il computo delle forze di interazione ed il modello di rappresentazione dell'acqua che si vorrà adottare. Il programma `pdb2gmx` del pacchetto GROMACS ci consente di fare questo e la nostra scelta, come detto in precedenza ed in analogia con le simulazioni di Maffeo [21], ricadrà sul f-f AMBER03 ed un modello di acqua a tre siti di interazione di tipo TIP3P.

Ora che la topologia è stata creata è necessario definire la dimensione del box all'interno del quale eseguiremo la simulazione. A questo punto occorre precisare che eseguiremo diversi tipi di simulazione a seconda delle caratteristiche che desideriamo investigare differenziandole

in due categorie:

- simulazioni a singola molecola (singleDNA), effettuate all'interno di box cubici a diverse dimensioni;
- simulazioni a 18 DNA (multiDNA), ottenute replicando un box tetragonale contenente una singola molecola tre volte in ciascuna direzione x ed y e due volte lungo z, ottenendo così sistemi formati da  $N_{DNA} = 3 \cdot 3 \cdot 2 = 18$ .

In tabella 4.2 troviamo le varie configurazioni trattate nella tesi, alle quali verranno assegnati dei nickname che ci consentiranno di non appesantire ulteriormente la scrittura.

nickname	box type	$N_{DNA}$	$l_x^{box}$	$l_y^{box}$	$l_z^{box}$	$V_{box}$
1.0x	cubic	1	6.078nm	6.078nm	6.078nm	224.49nm <sup>3</sup>
2.0x	cubic	1	7.658nm	7.658nm	7.658nm	449.04nm <sup>3</sup>
3.0x	cubic	1	8.778nm	8.778nm	8.778nm	676.29nm <sup>3</sup>
1DNA	tetragonal	1	2.659nm	2.710nm	5.082nm	36.608nm <sup>3</sup>
18DNA	tetragonal	18	7.976nm	8.129nm	10.163nm	658.94nm <sup>3</sup>

Tabella 4.2: In tabella sono riportate le tipologie di box considerate e le loro dimensioni. Per quanto riguarda i box tetragonali vengono forniti solamente i dati per la sequenza AAT, poiché i valori per le altre sequenze differiscono di poco.

Una volta preparato il box contenente gli acidi nucleici da analizzare occorre riempirlo con molecole d'acqua che costituiranno il solvente all'interno del quale avviene la dinamica. Questa operazione è affidata al software `genbox` che si occupa di inserire casualmente molecole di  $H_2O$  in modo tale da riempire l'intero sistema. Anche se non strettamente necessario, decidiamo di effettuare una prima minimizzazione dell'energia potenziale del sistema affinché si rilassino eventuali posizionamenti delle molecole d'acqua in posizioni sfavorevoli.

Arrivati a questo punto è importante considerare che la presenza di 12 coppie di basi azotate legate tra loro da 22 gruppi fosforici determina una carica elettrica netta del sistema di  $-22e$ . E' necessario neutralizzare tale carica per non introdurre degli effetti legati ad interazioni elettrostatiche, a questo scopo perciò introduciamo 22 controioni di sodio  $Na^+$  mediante il programma `genion`, che sostituisce casualmente alcune molecole d'acqua con gli ioni desiderati posizionandole in siti energeticamente favorevoli.

In realtà nelle nostre simulazioni aggiungeremo ulteriormente sale, inserendo separatamente ioni di cloro  $Cl^-$  e di sodio  $Na^+$ , fino ad arrivare a concentrazioni  $1M$ . Questo accorgimento è particolarmente utile soprattutto nelle simulazioni a singola molecola per schermare elettrostaticamente il sistema limitando eventuali interazioni elettromagnetiche con se stesso dovute alle condizioni periodiche al bordo.

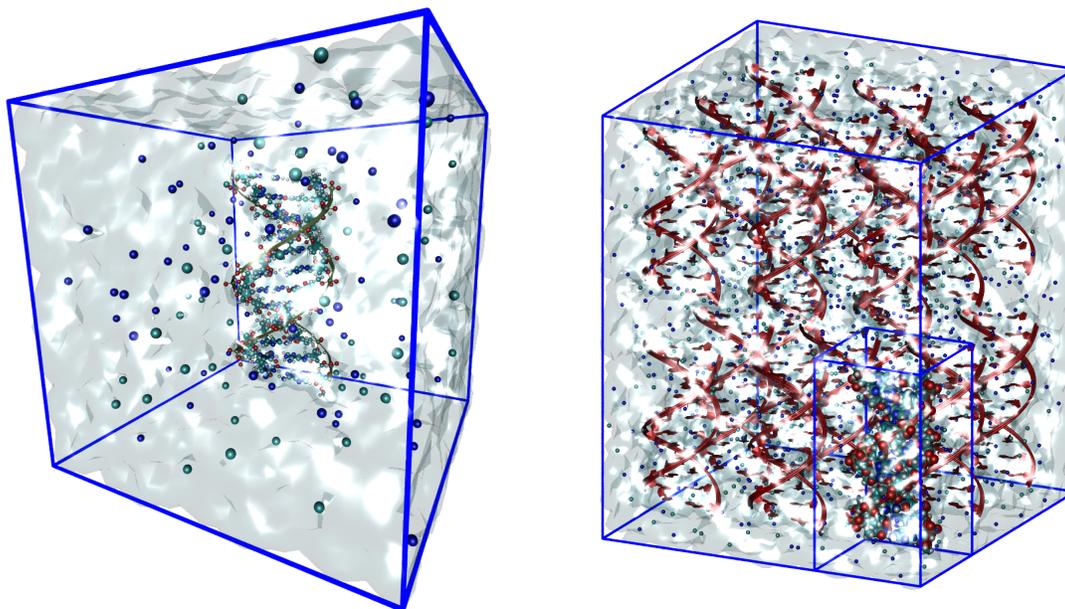


Figura 4.4: A sinistra rappresentazione di un singolo acido nucleico AAT all'interno di un box cubico idratato con ioni in concentrazione  $1M$  mentre a destra un sistema costituito da 18 molecole in un box tetragonale con la stessa concentrazione salina. Il DNA è rappresentato usando un tipo di visualizzazione “full-atom” chiamata CPK unita ad una “Ribbons” per evidenziare la doppia elica. L'acqua viene rappresentata solo per osservarne la presenza mediante una rappresentazione “QuickSURF” che ne evidenzia la superficie. Entrambi i rendering sono eseguiti mediante l'utilizzo del linguaggio POV-Ray 3.6 il primo usando l'opzione di visualizzazione “perspective” e l'altro la “orthographic”.

A questo punto i sistemi sono pronti per la simulazione previa ulteriore minimizzazione dell'energia potenziale del sistema.

## 4.5 Minimizzazione dell'energia potenziale

Prima di iniziare la simulazione vera e propria è necessario far rilassare il sistema per scongiurare eventuali posizionamenti degli atomi in posizioni sfavorevoli dal punto di vista energetico.

Per fare questo utilizzeremo dapprima lo strumento **grompp** che prendendo in input la configurazione finale di ciascun sistema, la topologia ed un file \*.mdp contenente tutti i parametri della simulazione, genera un file \*.tpr mediante il quale eseguire la simulazione vera e propria attraverso il comando **mdrun**. Per tutti i sistemi considerati è risultato appropriato utilizzare un numero di circa 150 steps di minimizzazione per ottenere un sistema energeticamente stabile.

Come detto in precedenza, eseguiremo ogni volta due minimizzazioni dell'energia seppur la prima non sia strettamente necessaria. E' fondamentale invece il processo di minimizzazione dell'energia dopo l'aggiunta degli ioni che, introdotti casualmente per sostituzione delle molecole d'acqua, con buona probabilità saranno collocati in zone ad alta energia rendendo il sistema non perfettamente stabile e poco realistico.

Una volta effettuato il processo di minimizzazione per tutti i sistemi presi in esame, notiamo che le energie dei sistemi contenenti un singolo acido nucleico ( $E_{1DNA}$ ) ed anche quelli multi-molecolari ( $E_{18DNA}$ ) si attestano attorno a valori di:

$$E_{1DNA} = (3.95 \pm 0.02) \cdot 10^5 \frac{kJ}{mol} \quad (4.30)$$

$$E_{18DNA} = (1.27 \pm 0.02) \cdot 10^6 \frac{kJ}{mol} \quad (4.31)$$

sostanzialmente indipendenti dal tipo di sequenza contenuta all'interno del box.

Prima di lanciare qualsiasi simulazione è necessario controllare che tali valori di energia raggiungano un minimo, altrimenti rischiamo di far partire la simulazione da una condizione di equilibrio instabile che porterebbe all'inutilizzabilità dei dati ottenuti.

A titolo di esempio vengono riportati due grafici che rappresentano la diminuzione dell'energia potenziale e del potenziale di Lennard-Jones di configurazioni AAT a singolo ed a multiplo doppio filamento di DNA. Questi diagrammi ci permettono di notare quanto sia importante la seconda minimizzazione effettuata, poiché l'introduzione casuale di ioni all'interno del solvente genera posizionamenti ad alta energia che vanno necessariamente rilassati.

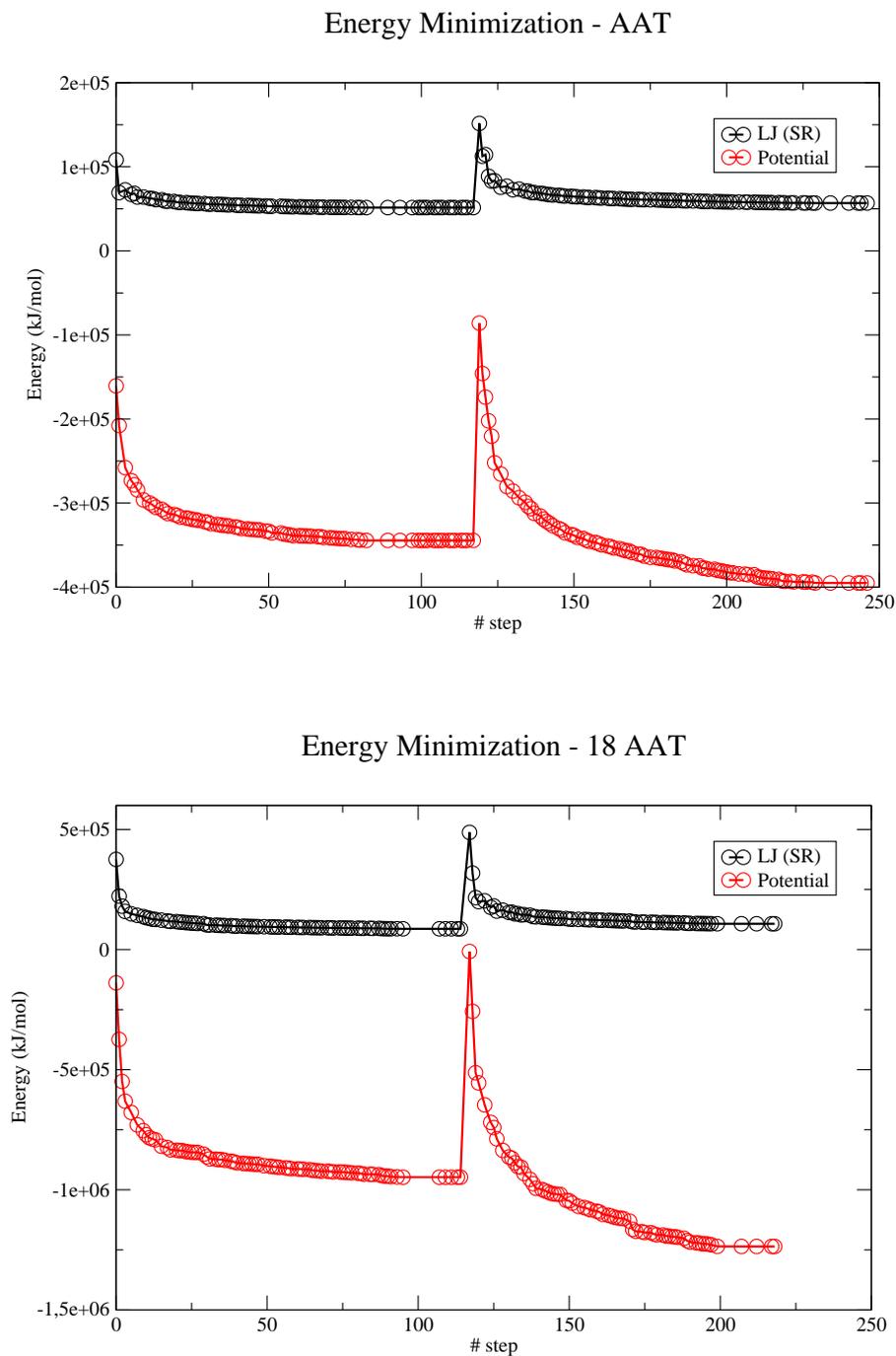


Figura 4.5: Processi di minimizzazione dell'energia potenziale del sistema AAT. Risulta particolarmente evidente la discontinuità energetica introdotta dall'aggiunta degli ioni carichi all'interno del sistema.

## Capitolo 5

# Parallelizzazione su SuperCalcolatori

### 5.1 Parallel-Computing con MPI

In tutte le versioni di GROMACS utilizzate per le simulazioni in questa tesi, è disponibile il supporto alle librerie MPI (Message Passing Interface) che consentono di ripartire il calcolo delle simulazioni ad un gran numero di processori distribuiti su di uno stesso cluster.

Il solo software che supporta tale architettura è l'`mdrun` perciò andremo ad eseguire le nostre simulazioni su cluster mentre l'analisi, utilizzando gli altri software del pacchetto, potrà essere fatta su di un normale pc domestico. A partire dalla versione 4.5, grazie all'introduzione dei "threads", `mdrun` supporta la parallelizzazione automatica su processori appartenenti ad uno stesso nodo senza esplicita richiesta delle librerie MPI.

La logica con cui viene implementato tale processo di parallelizzazione è piuttosto semplice, le librerie MPI infatti fanno uso dei cosiddetti "collaboratori", entità logiche che svolgono parallelamente determinati calcoli e che hanno la possibilità di comunicare tra loro attraverso il "message-passing", ovvero lo scambio di messaggi tra loro stessi.

Idealmente all'aumentare del numero  $P$  di collaboratori il tempo  $t$  di esecuzione del programma scalerà in modo lineare come pure la mole di dati  $A$  che dovrà gestire:

$$t^P = \frac{t^1}{P} \quad A^P = \frac{A^1}{P}. \quad (5.1)$$

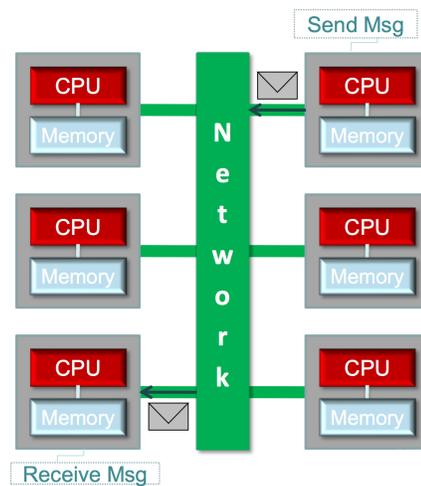


Figura 5.1: Scambio di messaggi tra varie diverse CPU attraverso il protocollo MPI.

Vedremo che questo non sarà mai così a causa di molti fattori ed in generale occorrerà trovare il giusto compromesso tra prestazioni e velocità di esecuzione nella scelta del numero di processori da utilizzare. Uno dei principali colli di bottiglia della parallelizzazione sarà la sincronizzazione della scrittura su disco dei dati processati provenienti da ogni CPU, unita alla frequente comunicazione tra i vari nodi di calcolo specialmente per quanto riguarda il calcolo delle interazioni a lunga distanza.

Anche in questo caso `mdrun` offre diverse opzioni per l'assegnazione di domini a ciascun processore utilizzato:

- “Particle Decomposition” (P-D), il tipo più semplice di decomposizione di un sistema atomico in cui gruppi di particelle vengono assegnati a ciascuno dei processori utilizzati. Per il calcolo delle forze questo metodo necessita che ogni processore conosca la posizione di almeno la metà delle particelle del sistema e per un gran numero di processori utilizzati dovranno quindi essere comunicate circa  $N \times N/2$  coordinate ad ogni passo di integrazione. Per colpa di questa dipendenza quadratica tale decomposizione non scala linearmente con il numero di processori a disposizione.
- “Domain Decomposition” (D-D), disponibile solo dalla versione 4.0, consiste nel suddividere il sistema in regioni spaziali a forma di parallelepipedo assegnandole a ciascun processore. Tali domini sono indipendenti dalle particelle al loro interno che quindi durante la simulazione possono passare dall'uno all'altro liberamente non essendo vincolate ad un dato processore. Questo metodo risulta più efficace del precedente poiché molte interazioni nella dinamica molecolare sono a carattere locale.

La nostra scelta ricadrà ovviamente sulla seconda tecnica di decomposizione spaziale, la più veloce nel nostro caso, lasciando al software la libertà di scegliere la migliore tipologia di suddivisione in blocchi a seconda del numero di processori a disposizione. Notiamo inoltre che dato che la trattazione delle interazioni elettrostatiche a lungo raggio è effettuata mediante l'algoritmo PME, sarà necessario suddividere<sup>1</sup> attraverso la D-D anche lo spazio reciproco in sotto-volumi da assegnare anch'essi a ciascun processore. Perciò il numero totale  $N$  di processori a disposizione verrà ripartito tra spazio diretto e reciproco di modo tale che:

$$N = N^{spatial} + N^{PME} \quad (5.2)$$

in cui  $N^{spatial} = N_x \times N_y \times N_z$  indica come suddividere i lati del box nelle tre direzioni spaziali e  $N^{PME} = N_{K_x} \times N_{K_y} \times N_{K_z}$  la suddivisione del volume in trasformata di Fourier.

---

<sup>1</sup>Anche in questo caso la suddivisione viene lasciata fare al programma di dinamica molecolare.

## 5.2 Parallelizzazione con Cluster Matrix

A causa della grande potenza di calcolo richiesta per la simulazione, per produrre simulazioni della maggiore durata possibile, avremo bisogno di far girare il programma su di un cluster di calcolatori.

Abbiamo a disposizione l'ormai datato cluster Matrix gestito dal consorzio nazionale di calcolo CASPUR, con sistema operativo Linux basato su architettura AMD Opteron Quad-Core con frequenza di clock di  $2.1GHz$  ed equipaggiati con  $16Gb$  di RAM e connessione InfiniBand. Il cluster è formato da 2 nodi che gestiscono il servizio di front-end per gli utenti, da 4 nodi interattivi per l'esecuzione di programmi in tempo reale che non richiedono grandi parallelizzazioni, e di 320 nodi di calcolo vero e proprio. Matrix ci consente di parallelizzare l'esecuzione di programmi su un massimo di 64 nodi, ciascuno dei quali equipaggiato da 2 Opteron, per un totale di 8 cores per nodo, che determinano delle performance di calcolo di  $22Tflop$ .

Ogni utente ha poi a disposizione una directory `/home` su cui poter salvare 200MB di dati, ed una directory `/work` da 10GB di spazio disponibile, è cruciale quindi eseguire le simulazioni sulla directory di lavoro, a causa del grande spazio di archiviazione richiesto (circa 3GB di dati per una simulazione di 300ns).



Figura 5.2: Una foto del cluster Matrix a sinistra, mentre sulla destra particolare dell'architettura Quad-Core dei processori Opteron che equipaggiano il Matrix.

Il sistema che gestisce le code tra gli utenti delle varie esecuzioni è TORQUE (Terascale Open-Source Resource and QUEUE Manager) coadiuvato da Moab, un software di management delle esecuzioni di programmi su cluster mentre il bilancio delle ore di calcolo a disposizione è affidato al software open source Gold. Per sfruttare il cluster Matrix avevamo un “grant” che ci mette a disposizione 200000 ore di esecuzione, sufficienti a studiare la conformazione di 3 tipi diversi di dodecameri di DNA. La durata massima del run che possiamo eseguire (“walltime”) dipende dal numero di nodi richiesti: per un massimo di 16 nodi avremo a disposizione massimo

40 ore di esecuzione, mentre per 64 nodi solamente 10 ore di esecuzione continuativa. Inoltre è disponibile anche una coda speciale di “debug”, per effettuare test veloci per settaggi ed ottimizzazioni del sistema, il cui “walltime” è limitato a soli 30 minuti.

Per ottimizzare la simulazione occorre studiare come si comporta il cluster al variare dei nodi a disposizione, infatti aumentare a dismisura il numero di processi in cui suddividere il sistema potrebbe essere dispendioso in termini di tempo di esecuzione del run.

Per avere un termine di riferimento per le performance facciamo perciò girare la medesima simulazione, scelta ad-hoc per l’occasione<sup>2</sup>, su di un normale pc domestico e confrontiamo i tempi di esecuzione al variare dei cores usati sul cluster Matrix. Il numero di step settati è  $10^5$  e pc di riferimento, processore Intel i5 Quad-Core 2.7Ghz dotato di 4Gb di ram, otteniamo come tempo di riferimento il valore di  $t_{run} = 1293.11s$ .

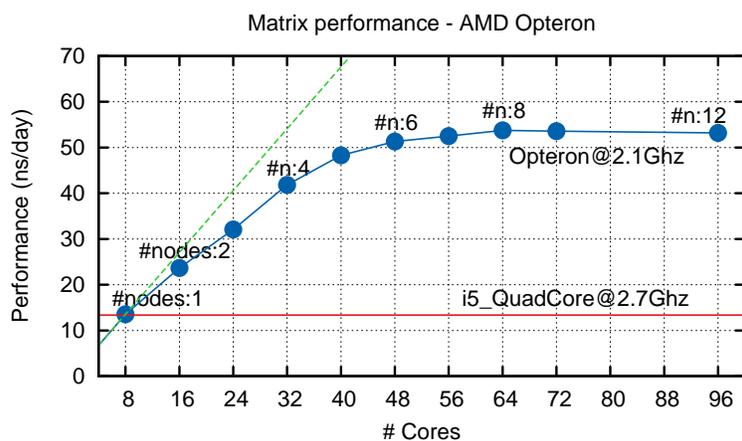


Figura 5.3: Andamento del tempo di esecuzione della simulazione in funzione del numero di cores utilizzati. In verde è riportato il risultato della simulazione effettuata su una normale cpu domestica.

Come prima cosa è stato lanciato un processo su singolo nodo sfruttando un solo core, in questo caso la simulazione ha richiesto un tempo secolare di  $t_{run} = 10909.09s$  che non è stato riportato in figura. I run successivi sono stati fatti al variare del numero dei nodi ed utilizzando il massimo numero di cores a disposizione su ciascun nodo (8 cores per nodo) per ridurre al minimo la presenza di “colli di bottiglia” nella trasmissione dati. Notiamo che già utilizzando 2 nodi il tempo di esecuzione si è più che dimezzato e le massime performance sono ottenute parallelizzando la simulazione in 64 processi, ovvero usando 8 nodi. Il sistema configurato in questo modo impiega  $t_{run}^{8 \times 8} = 266.51s$  per completare il run.

<sup>2</sup>Un sistema formato da un solo acido nucleico di tipo AAT in un volume 1x.

Aumentando ancora la parallelizzazione non aumentano i benefici in termini di prestazioni, infatti il sistema si riattesta su tempi  $t_{run} \simeq 320s$  comparabili ai valori ottenuti con 6 o 7 nodi. Questo è dovuto al fatto che GROMACS suddivide il sistema in un numero di parti uguali al numero di cores a disposizione e, mediante il bus Infiniband, ciascun processore comunica con gli altri parallelizzando il sistema. Perciò all'aumentare della parallelizzazione aumenta anche l'utilizzo del bus InfiniBand che ad un certo grado di parallelizzazione rallenta l'esecuzione del run, determinando un effetto peggiorativo per la tempistica di esecuzione.

Ovviamente, quindi, è il volume del sistema che influisce pesantemente sulla parallelizzazione e quindi questo studio andrà ripetuto nei casi in cui considereremo un volume differente.

Da ora in avanti utilizzeremo la parallelizzazione con 64 processi, eseguiti su 8 nodi dotati di 8 cores ciascuno, purtroppo però l'utilizzo del cluster Matrix sarà interrotto il 31 Dicembre 2012, data nella quale verrà spento e disassemblato in via definitiva, in seguito all'accorpamento dei tre maggiori consorzi di calcolo italiani (CASPUR, CILEA, CINECA) in un unico consorzio nazionale (CINECA).

### 5.3 Parallelizzazione con JUROPA/HPC-FF

La parte relativa allo studio sugli effetti di taglia del sistema la andremo ad effettuare su un più performante supercomputer in Germania. Lo Jülich Supercomputing Centre (JSC) dell'Institute for Advanced Simulation (IAS) con sede a Jülich, ci ha messo a disposizione l'utilizzo del suo moderno cluster JUROPA/HPC-FF: il nostro progetto sullo studio del ruolo della conformazione dei segmenti di DNA corti nella transizione di fase nematica ci ha permesso di vincere un "grant" da 100,000 ore di utilizzo del cluster.

Il cluster JUROPA/HPC-FF è diviso in due partizioni principali, la JSC con grandi risorse hardware, e la HPCFF dotata di circa la metà di potenza computazionale rispetto alla prima. Il "grant" vinto ci permette di usare la partizione JSC costituita da 2208 nodi (di cui 32 per l'interfaccia del sistema e 2176 di calcolo) mentre la HPCFF dispone di 1080 nodi (16 per l'interfaccia, 1064 per il calcolo) ovvero circa la metà delle risorse hardware della JSC.

Ciascun nodo di calcolo è dotato di due processori Intel Xeon X5570 (Nehalem-EP) Quad-Core, con frequenza di clock di  $2.93GHz$  e tecnologia SMT (Simultaneous Multithreading), per un totale di 8 cores equipaggiati con  $24GB$  di memoria RAM DDR3 a  $1066MHz$ . I nodi di calcolo comunicano tra loro mediante il bus InfiniBand Mellanox ConnectX QDR HCA (con un data rate massimo teorico di  $96Gbit/s$ ) che garantisce alte prestazioni in fase di parallelizzazione, permettendo ai 17664 cores del cluster JCS di raggiungere un picco di  $207Tflop$  di performance (mentre gli 8640 cores dell'HPCFF permettono performance di  $101Tflop$ ).

Ogni utente ha a disposizione  $3TB$  di memoria di archiviazione nella directory  $\$HOME$  ed altrettanti nella  $\$WORK$ , mentre una grande unità di storage permette agli utenti di salvare fino

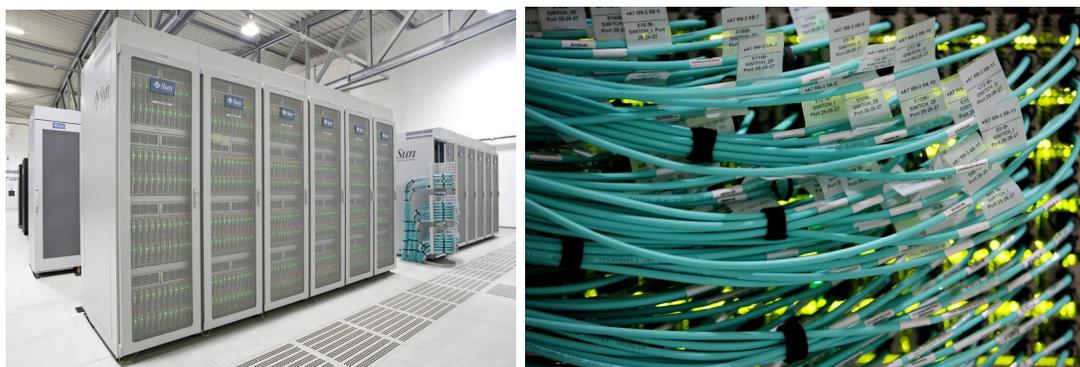


Figura 5.4: A sinistra il cluster Juropa JCS, mentre sullo sfondo è possibile notare l'HPCFF (rack neri). Nella foto a destra particolare della connessione InfiniBand nel cluster JCS.

a  $100TB$  di dati nella  $\$GPFSARCH$ . Non avremo quindi problemi di spazio, dato che i singoli run occupano al massimo  $\sim 3GB$  di dati per  $300ns$  di simulazione.

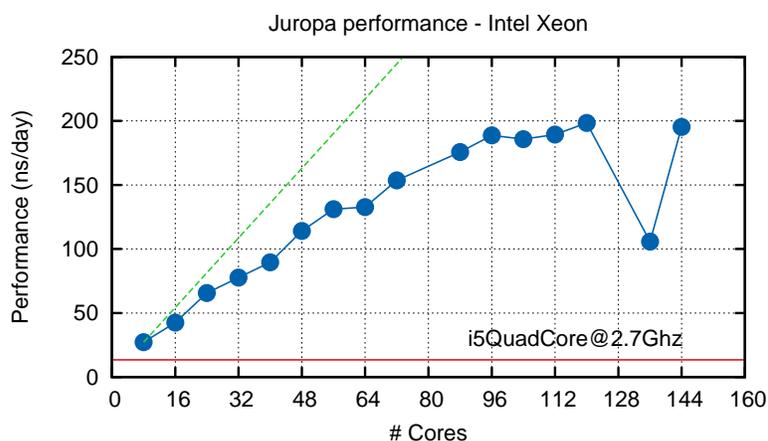


Figura 5.5: Andamento del tempo di esecuzione della simulazione in funzione del numero di cores utilizzati. In verde è riportato il risultato della simulazione effettuata su una normale CPU domestica.

In particolare è possibile eseguire run della lunghezza di 6 ore nella modalità interattiva (Interactive Jobs) su di un massimo di 8 nodi, mentre la lunghezza dei run è di 24 ore nella modalità Batch, la modalità di calcolo vera e propria, che permette di sfruttare un massimo di 1024 nodi per ciascun run.

Effettuando i test al variare del numero di cores utilizzati apprezziamo come il SPC tedesco scali in modo molto lineare fino a circa 128 processori dopodichè la velocità di esecuzione satura e non otteniamo più benefici dall'aumento del numero. Decidiamo perciò di effettuare le nostre

simulazioni usando 96 cores, un giusto compromesso tra efficienza e potenza di calcolo. Inoltre, siccome quasi tutte le simulazioni a molti DNA verranno eseguite su Juropa ripeteremo lo studio fatto sin qui anche nel nuovo caso, ottenendo che il numero ideale di processori con cui far girare le simulazioni risulta essere pari a 128.

## 5.4 Parallelizzazione con Fermi BlueGene/Q

Dopo lo spegnimento del cluster Matrix e della fusione dei tre consorzi di calcolo nazionali abbiamo la possibilità di usare il supercomputer Fermi basato su architettura IBM BlueGene/Q e dotato perciò di moderni processori IBM PowerA2 con frequenza di clock di  $1.6GHz$  da 18 cores ciascuno: 16 cores adibiti al calcolo vero e proprio, 1 utilizzato per le funzionalità dei servizi del sistema operativo (come ad esempio della gestione dell'I/O asincrono) ed un altro utilizzato solamente in caso di danneggiamento di uno degli altri. Tali processori sono configurati su 10240 nodi di calcolo da 16 cores ciascuno ed equipaggiati con  $16Gb$  di RAM. In totale il BG/Q dispone di un impressionante numero di cores computazionali, per la precisione 163.840, che determinano performance di calcolo eccellenti pari a  $2.1PFlops$  e che permettono al Fermi di raggiungere la 9° posizione nel ranking mondiale per potenza di calcolo<sup>3</sup>.

In questo caso la connessione interna è affidata ad un'interfaccia Torus ad elevatissime prestazioni, connessione che ben si presta ad una massiva parallelizzazione dei programmi in esecuzione.

Lo scheduler nativo per le esecuzioni dei programmi batch è il LoadLeveler (LL), che provvede a gestire i jobs nel sistema allocando una partizione per ciascun nodo, gestendo anche l'output del job ed eventuali file di errore. Il Fermi ci mette a disposizione diverse partizioni di storage dei dati, la principale è la cartella locale `$HOME` nella quale possiamo salvare  $4GB$  di dati permanenti ed a backup periodico, l'altra partizione è locata in `$CINECA_DATA` in cui abbiamo a disposizione  $100GB$  di memoria di archiviazione permanente ma della quale non viene fatto alcun backup, inoltre tale memoria è condivisa ed

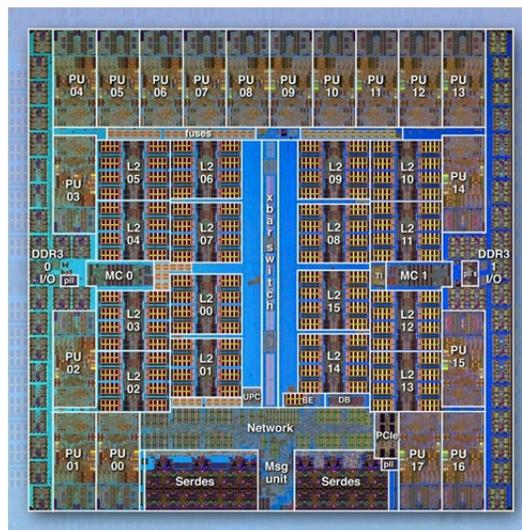


Figura 5.6: Particolare del complesso schema di fabbricazione del processore IBM PowerA2. Notiamo la caratteristica composizione a 18 cores e relative cache dedicate.

<sup>3</sup>Al momento in cui stiamo scrivendo la classifica è stata rivista ed il Fermi è sceso in 12° posizione, soppiantato tra l'altro dal nuovo cluster di Jülich JUQUEEN con architettura BlueGene/Q.

accessibile a tutti gli utenti. L'ultima parte di memoria è la cartella `$CINECA_SCRATCH`, cartella locale e temporanea (i dati vengono cancellati dopo 30 giorni di mancato utilizzo) che viene montata direttamente sul nodo di esecuzione, per questo i programmi che necessitano di un batch-script devono essere salvati qui, infatti sarà la directory di lavoro nella quale faremo girare le nostre simulazioni. Anche in questo caso non abbiamo problemi di storage poiché non c'è limite alla dimensione di tale directory. Il grant che abbiamo a disposizione ci consente un utilizzo di 200000 ore di run, successivamente estese a 300000 per far fronte alle nuove esigenze di calcolo che spiegheremo fra poco.

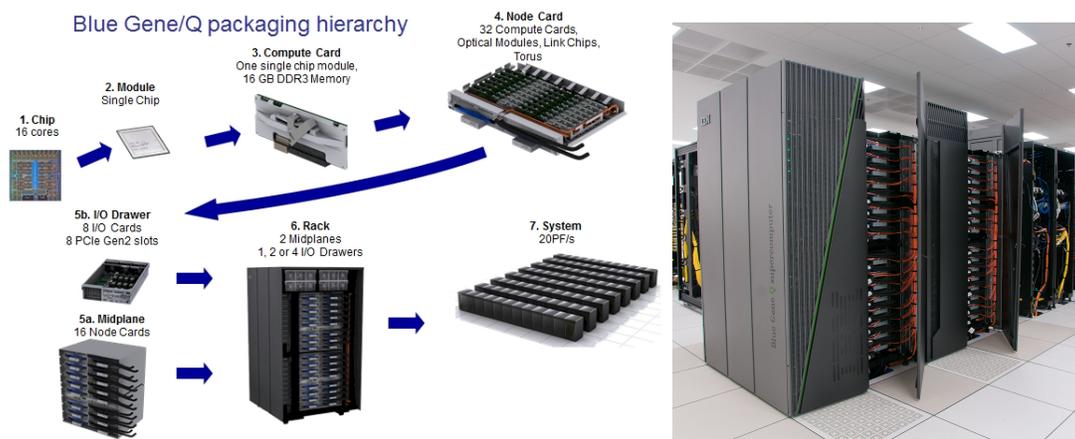


Figura 5.7: Schema della gerarchia costruttiva del supercomputer Fermi. A destra particolare del rack e delle interconnessioni tra i nodi di calcolo.

Per sfruttare a pieno le performance del BG/Q dobbiamo riconfigurare il sistema adattandolo ad un'alta parallelizzazione tenendo conto del lavoro [43] fatto su sistemi analoghi.

Tale sistema ammette massima decomposizione a 640 domini nello spazio reale, che verranno quindi assegnati, dove possibile, a ciascun core. Da questo valore in poi i processori in sovrannumero vengono assegnati alla computazione del PME, settando esplicitamente il comando `-npme <#cores>` in fase di esecuzione.

Nonostante il supercomputer in dotazione sia stato riconosciuto al 9° posto nel ranking mondiale per potenza di calcolo, riusciamo ad ottenere prestazioni a malapena comparabili con il ben più datato Cluster Matrix, come mostrato in figura, ed utilizzando un numero di cores di un'ordine di grandezza superiore!

Notiamo che la scala temporale si attesta grossomodo sullo stesso ordine di grandezza del cluster Matrix, ma ottenendo performance con efficienza dimezzata. Probabilmente questo è dovuto al fatto che il nostro sistema è davvero molto piccolo, perciò non riusciamo a sfruttare tutta la potenza di parallelizzazione offerta dal BG/Q. In particolare nel nostro caso specifico, influisce in modo negativamente decisivo sulle performance la ridotta potenza di calcolo del

singolo core,  $1.6GHz$  non riescono sicuramente a tenere il passo con i ben più veloci processori esaminati in precedenza.

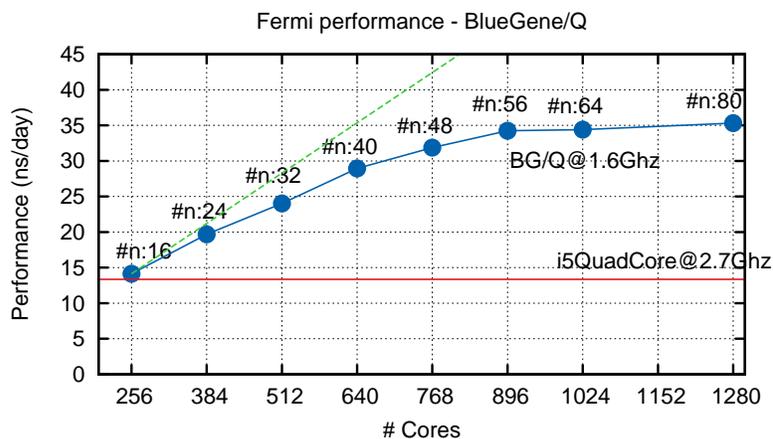


Figura 5.8: Andamento del tempo di esecuzione della simulazione in funzione del numero di cores utilizzati. In verde è riportato il risultato della simulazione effettuata su una normale cpu domestica.

Deduciamo perciò che tale architettura non si presta bene alla simulazione di dinamica molecolare con ridotto numero di atomi (20k atomi presenti nel nostro sistema) anche perché utilizzando un numero di core così elevato esauriamo davvero rapidamente le ore di run a nostra disposizione<sup>4</sup>.

<sup>4</sup>Peccato non aver potuto provare il BGQ su un sistema più grande contenente molti DNA, l'idea di simulare questo sistema ci è venuta dopo aver terminato le ore a disposizione sul Fermi.

## Capitolo 6

# Parallelizzazione su GPU

### 6.1 Parallel-Computing con CUDA

La continua crescita di potenza e la sofisticazione nello sviluppo tecnologico delle recenti schede video, stimolata dal settore videoludico e dalla sempre più grande richiesta grafica degli ultimi sistemi operativi, ha fatto sì che tali dispositivi passassero da un ruolo di nicchia ad occupare un posto di centrale importanza nell'hardware di qualsiasi PC. L'ingegnerizzazione delle GPU (Graphic Processing Units) ha seguito un percorso parallelo e piuttosto indipendente rispetto allo sviluppo delle CPU, spingendo gli studiosi a progettare architetture che dovessero affrontare in tempo reale la risoluzione di problemi ad alto indice di parallelizzazione (come ad esempio il "rendering" di una scena di un videogames).

In particolare nVidia Corporation, una delle aziende leader nel settore, ha tentato di unificare l'architettura hardware delle sue GPU dando vita al progetto CUDA (acronimo di Compute Unified Device Architecture) che propone un nuovo ed interessante approccio all'elaborazione parallela. Attraverso un ambiente di sviluppo apposito nVidia ha reso disponibile ai programmatori un nuovo strumento di programmazione che ben si presta all'utilizzo in ambito scientifico e che prende il nome di CUDA-C (anche se spesso per brevità viene omessa l'appendice -C), un paradigma di programmazione ad alto livello basato sull'ormai consolidato ed onnipresente linguaggio C anche se sono presenti altre estensioni basate su Python, Fortran, Java e MATLAB.

Dato il notevole successo dell'iniziativa proposta in esclusiva da nVidia (infatti il codice di sviluppo di CUDA è freeware ma non è uno standard aperto), è stato sviluppato anche un altro linguaggio di programmazione libero, chiamato OpenCL, in analogia ai già esistenti OpenGL ed OpenAL, capace di interfacciarsi con tutte le altre schede grafiche disponibili sul mercato, tra cui su tutte le schede AMD ATI che sinora erano state tagliate fuori da questo settore. OpenCL, sviluppato e supportato inizialmente da Apple ma ora in fase di beta testing su Linux e Windows, è anch'esso un linguaggio ad alto livello basato sul C-99 anche se più

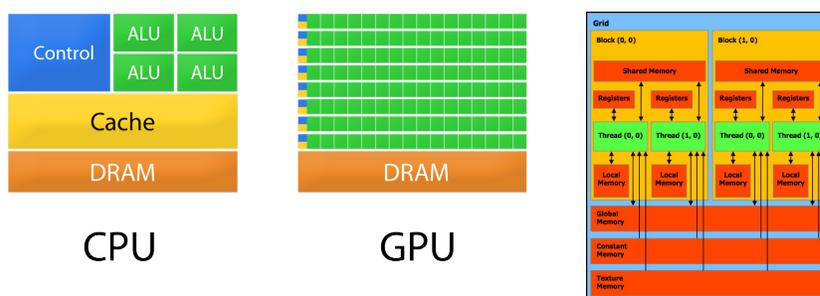


Figura 6.1: Nella schema a sinistra vengono mostrate le notevoli differenze progettuali tra CPU e GPU, ciascun quadrato verde rappresenta il singolo core addetto all'elaborazione dati. A destra invece lo schema griglia-blocchi-threads col quale viene implementata la programmazione CUDA.

complicato nella sintassi rispetto a CUDA e più pesante nella scrittura del codice, un piccolo prezzo da pagare per il supporto ad architetture diversificate.

Nel momento in cui scriviamo è appena stata rilasciata la nuova release di GROMACS 4.6 con il supporto al linguaggio CUDA, che quindi apre le porte della dinamica molecolare alla straordinaria potenza di calcolo offerta delle GPU. Purtroppo, per ora, non è ancora disponibile una release analoga col supporto ad OpenCL anche se presto verrà rilasciata.

Analizziamo perciò come questo nuovo strumento possa essere utilizzato in maniera efficiente nell'ambito del parallel computing e più in particolare nella simulazione di dinamica molecolare. L'approccio alla programmazione su schede grafiche viene chiamato in generale GPGPU (General-Purpose computing on Graphics Processing Units) e si propone come filosofia di programmazione un paradigma del tipo "Single-Instruction Multiple-Thread", in cui la vera efficienza computazionale viene raggiunta proprio eseguendo uno stesso tipo di operazione su un gran numero di processi identici.

La logica con la quale viene progettata una CPU è ottimizzata per l'esecuzione di un codice sostanzialmente seriale ed il numero di core presenti è limitato a poche unità di calcolo ALU (Arithmetic Logic Unit), le quali sono controllate da una sofisticata unità di controllo CU (Control Unit) che deve essere in grado di monitorare una grandissima quantità di comandi diversi da eseguire. Alla luce di questo fatto per velocizzare ulteriormente le operazioni, questa architettura è assistita da una grande memoria cache, per ridurre al massimo le latenze con cui la CPU accede alla memoria centrale. Tali caratteristiche non si prestano bene ad una esecuzione parallela, poiché la CPU punta tutto sulla velocità di computazione del singolo programma piuttosto che sulla suddivisione del problema sfruttando la parallelizzazione su ciascun cores.

L'architettura di una GPU ben si presta ad una parallelizzazione massiva ed infatti una singola unità integra un gran numero di cores denominati CUDA-Core, dell'ordine delle centinaia di unità su ogni scheda grafica e dotate ciascuna di una propria ALU, anche se la presenza

di unità di controllo è minimizzata proprio per il fatto che un gran numero di core eseguono la stessa operazione in parallelo.

Ma d'altro canto è pur vero che una GPU non può girare senza una CPU di supporto, quindi in generale i programmi in CUDA verranno di fatto separati in due sezioni: la parte seriale è affidata all'esecuzione su CPU mentre la parte fortemente parallelizzata verrà fatta girare sulla GPU. Questo può essere un collo di bottiglia nella computazione e dovremmo tenerne conto durante l'esecuzione ed inoltre, in generale, un altro rallentamento può essere dovuto al trasferimento di dati tra le due memorie.

Un programma che gira su una GPU viene chiamato kernel, ed è eseguito come una griglia ("grid") organizzata in blocchi ("blocks") contenenti tutti lo stesso numero di "threads" paralleli. Il singolo thread è quindi l'istruzione base che vuole essere parallelizzata, cioè eseguita contemporaneamente su tutti i CUDA-Core. Ciò non implica che il numero di threads del kernel mappi esattamente il numero di core della GPU, essi infatti verranno eseguiti a scaglioni occupando tutti i cores disponibili finché tutti i threads non avranno processato le operazioni richieste. A livello hardware l'esecuzione di uno o più blocchi di threads viene chiamata "Streaming Multiprocessor" (SM) mentre l'esecuzione del singolo thread su ciascun core viene chiamata "Streaming Processor" (SP).

La dimensione dei blocchi e delle griglie, gestite entrambe come array tridimensionali, dipende sostanzialmente solo dalla versione del software CUDA a disposizione, mentre ovviamente le caratteristiche hardware generali vengono contraddistinte in diverse classi, identificate dalla sigla "Compute Capability". Questa logica di mappatura tra software ed hardware è davvero flessibile e rende la programmazione in CUDA indipendente dalle specifiche tecniche della GPU su cui andiamo ad eseguire il programma.

## 6.2 Parallelizzazione su nVidia Tesla

Sebbene sia possibile far girare la nostra simulazione su qualsiasi tipo di scheda grafica nVidia, all'interno del cluster PATCHY dell'università La Sapienza di Roma abbiamo a disposizione un nodo di calcolo dedicato, costituito da 3 GPU nVidia Tesla due delle quali sono il modello

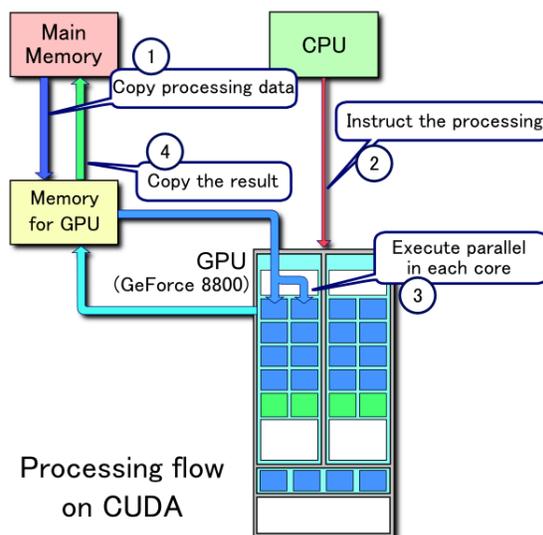


Figura 6.2: Schema che mostra il flusso di dati tra le componenti hardware durante l'esecuzione di un programma scritto in CUDA.

M2075 e l'altra modello M2070. Dal punto di vista hardware e prestazionale le due schede non differiscono di molto: entrambe dotate di 448 CUDA-Core con frequenza di clock di 1.15GHz e 6GB di RAM GDDR5 con una larghezza di banda di 150GB/sec, differiscono solamente per alcuni dettagli tecnici nella disposizione dei componenti hardware sulla scheda e dalla sigla che identifica il tipo di chip, T20 per la M2070 e T20A per la M2075. Anche dal punto di vista pratico (facendo eseguire dei run di prova selezionando ciascuna scheda) non troviamo differenze in termini di tempi di esecuzione, perciò le considereremo del tutto equivalenti.

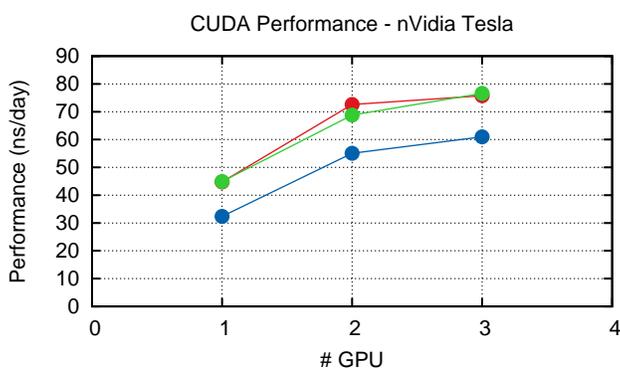


Figura 6.3: Performance dell'esecuzione delle simulazioni su GPU nVidia in funzione del numero delle schede video utilizzate.

Per poter iniziare con la simulazione occorre settare dei parametri particolari per la simulazione su GPU, in particolare la sezione che gestisce il “cut-off” v  settata usando il metodo di Verlet, l'unico finora disponibile per la gestione del troncamento dell'interazione su schede video.

Dai risultati ottenuti   chiara l'efficienza prestazionale del calcolo su GPU nonostante il supporto di GROMACS a tale tecnologia sia stato appena introdotto. Utilizzando sole tre GPU superiamo di gran lunga le prestazioni ottenute dal cluster Fermi usando migliaia di cores, attestandosi a produzioni di quasi  $80ns/day$  per un sistema a singolo DNA contenente circa  $20katomi$ . Abbiamo provato inoltre a far girare simulazioni ben pi  pesanti usando sistemi contenenti  $60katomi$  ottenendo velocit  di esecuzione pari a  $27ns/day$ , ovvero tre volte pi  lento della simulazione con un terzo delle particelle, dato che mostra l'eccellente scalabilit  della computazione GPGPU.

# Capitolo 7

## Analisi dati

### 7.1 Simulazioni di DNA in soluzione acquosa

Attraverso l'utilizzo degli SPC citati in precedenza saremo in grado di esaminare complessivamente 7 simulazioni a singolo acido nucleico e 3 contenenti 18 doppie eliche corte di DNA.

Prima di passare all'analisi dati vera e propria è necessario controllare che tutte le variabili termodinamiche del sistema si mantengano all'equilibrio lungo tutta la durata della simulazione. Scegliendo come insieme statistico l'NPT in cui il numero di particelle, la temperatura e la pressione siano controllate dall'esterno, occorrerà controllare per ciascuna delle simulazioni che le fluttuazioni delle variabili termodinamiche siano comparabili con gli esperimenti reali. In realtà avevamo visto come l'utilizzo degli algoritmi di "coupling" per pressione e temperatura consentissero la fluttuazione controllata delle due grandezze, non ponendoci esattamente in un ensemble NPT. Questo non cambia il nocciolo del discorso, anzi sarà necessario tenere sotto controllo un numero più grande di parametri.

Ovviamente il numero di particelle del sistema è fissato a priori durante la preparazione del box con successiva introduzione di acqua e non varierà mai durante l'evoluzione del sistema. In generale i sistemi che andremo a simulare sono composti da  $20 \div 60k$  atomi mentre i test per la scelta del numero ottimale di processori da utilizzare sono stati fatti con sistemi da  $20k$  atomi.

Nickname	$N_{atoms}$	$N_{atoms}^{water}$ (single DNA 1.0x)	$N_{atoms}^{water}$ (single DNA 2.0x)	$N_{atoms}^{water}$ (single DNA 3.0x)	$N_{atoms}^{water}$ (multiple DNA)
AAT	764	21915	40917	66417	62343
AAC	762	21910	40924	66415	62256
ACC	760	22441	40928	66413	62214

Tabella 7.1: Numero di atomi che compongono il sistema, sia nel caso del singolo DNA immerso in soluzione, sia contando esplicitamente tutti gli atomi, solvente e ioni compresi.

Nonostante l'aumento del numero delle interazioni possibili all'aumentare del numero di atomi, abbiamo notato un riscaldamento grossomodo lineare da parte dei tempi di esecuzione richiesti al sistema.

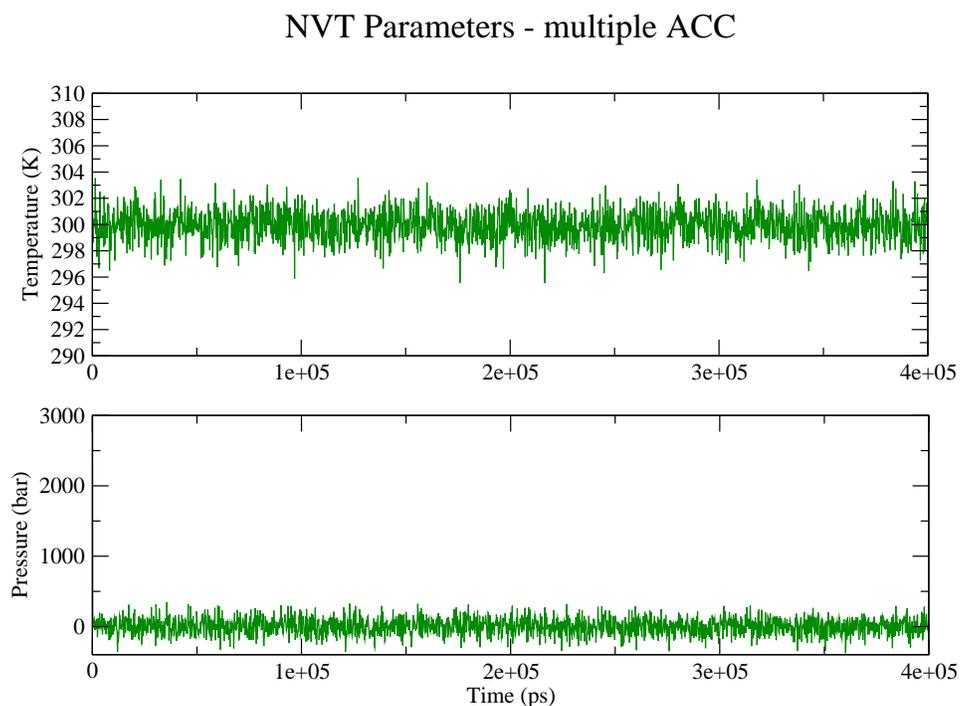


Figura 7.1: Andamento delle variabili termodinamiche nell'insieme NPT nel caso di un sistema formato da 18 sequenze di tipo ACC. Gli algoritmi di “coupling” per temperatura e pressione portano ad un corretto mantenimento dei valori settati.

I parametri più importanti da controllare saranno quelli legati alle variabili di pressione e temperatura e quelli legati al calcolo esplicito dell'energia totale del sistema. A titolo esplicativo mostriamo quelli per il sistema costituito da 18 sequenze ACC anche il controllo di queste variabili verrà effettuato per ciascuna simulazione. Attraverso i programmi `eneconv` ed `g_energy` sarà possibile analizzare tutte le variabili termodinamiche dei sistemi contenute nei file \*.edr ottenuti dopo il completamento della simulazione.

In generale otteniamo che per i valori dei parametri termodinamici si attestano attorno a:

$$T_{system} = (299.929 \pm 0.002) K \quad (7.1)$$

$$P_{system} = (1.031 \pm 0.038) bar \quad (7.2)$$

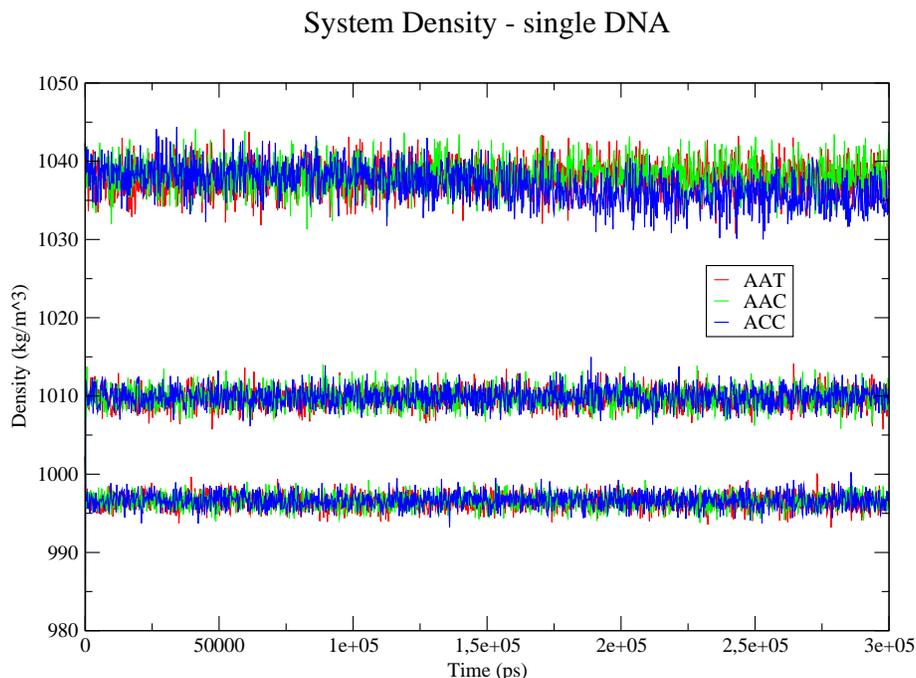


Figura 7.2: Andamento dei valori della densità del sistema contenente un singolo doppio filamento di DNA. Notiamo che non ci sono differenze sostanziali tra i tre sistemi.

$$E_{system}^{1DNA} = - (4.096 \pm 0.003) \cdot 10^5 \frac{kJ}{mol} \quad (7.3)$$

$$E_{system}^{18DNA} = - (1.505 \pm 0.005) \cdot 10^6 \frac{kJ}{mol} \quad (7.4)$$

come da valori impostati nel setup delle simulazioni. Questo mostra non solo che il termostato “v-rescaling” per le temperature e l’algoritmo di Berendsen per le pressioni mantengono le simulazioni a valori termodinamici comparabili con quelli di un esperimento di laboratorio controllando le fluttuazioni in modo efficiente, ma anche come la scelta del passo di integrazione sia consistente con una fluttuazione massima di tali grandezze del 3%.

Oltre alle variabili d’ensemble è interessante tenere sotto controllo la densità  $\rho$  del sistema, per vedere come essa si comporta durante l’evoluzione temporale. Nei casi in cui il sistema contiene una singola sequenza osserviamo che la densità è grossomodo stabile per tutte le dimensioni del box prese in esame ed inoltre non riscontriamo significative differenze al variare dell’acido nucleico.

Completamente differente è la situazione per quanto riguarda il più realistico sistema costituito da 18 DNA: la densità del sistema risulta criticamente dipendente dalla sequenza presa in esame e dai grafici si evince chiaramente che il sistema costituito da acidi nucleici di tipo

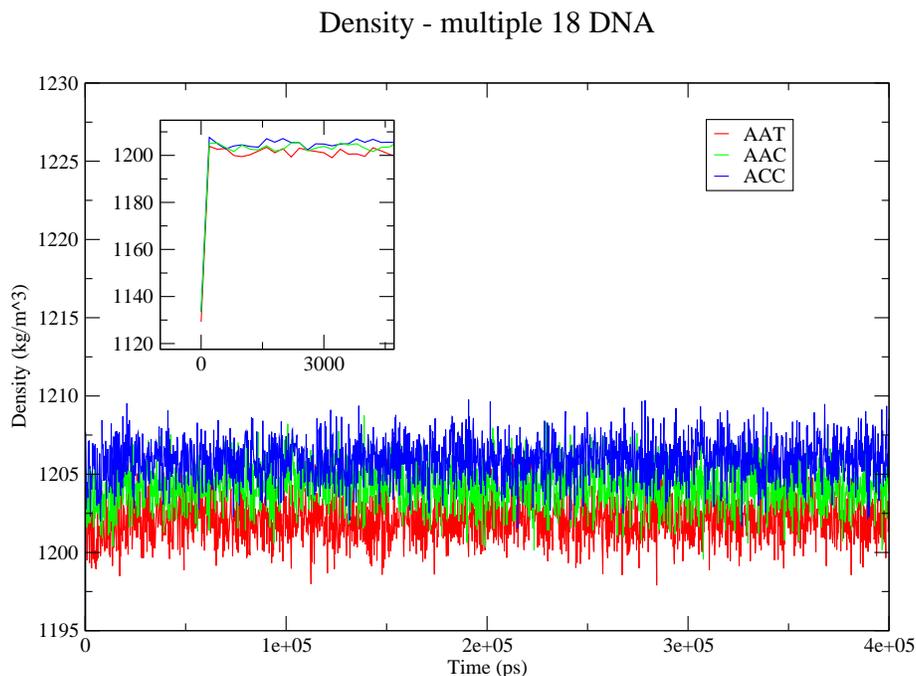


Figura 7.3: Fluttuazioni delle densità nei tre sistemi composti da 18 DNA ciascuno. Risulta particolarmente evidente come i valori di  $\rho$  si attestino rapidamente ai valori di equilibrio (da una stima grafica dopo 20ns circa il sistema ha raggiunto la densità media).

ACC, che secondo Nakata e Zanchetta [16] dovrebbe transire a concentrazioni maggiori, si stabilizza a densità maggiori. Ora, supponendo che le tre tipologie di sequenze abbiano lo stesso peso molecolare, questo comportamento ci fornisce un primo indicatore del fatto che la ACC possa essere la più piegata tra le tre sequenze prese in esame poiché una struttura piegata è geometricamente più compatta, quindi a densità maggiore, della sua controparte rettilinea.

Le medie delle tre densità nel caso del sistema multiDNA lungo tutta la durata della simulazione, si attestano a valori di circa:

$$\rho_{AAT}^{system} = (1202.02 \pm 2.10) \frac{kg}{m^3} \quad (7.5)$$

$$\rho_{AAC}^{system} = (1204.22 \pm 2.06) \frac{kg}{m^3} \quad (7.6)$$

$$\rho_{ACC}^{system} = (1205.86 \pm 2.07) \frac{kg}{m^3} \quad (7.7)$$

determinando cioè una variazione massima dello 0.3% nella densità totale del sistema.

Inoltre l'evidenza della stabilizzazione a densità differenti nel sistema a DNA multipli ci fa dubitare del fatto che le simulazioni a singolo filamento ci forniscano, nel seguito, dei dati

interessanti per quanto riguarda il comportamento medio dei sistemi simulati.

## 7.2 Equilibratura della struttura terziaria dell'acido nucleico

Abbiamo visto che le proprietà termodinamiche elementari del sistema sono ben conservate durante tutte le simulazioni, è lecito perciò effettuare un'analisi più accurata della struttura del DNA in soluzione. Prima di trarre qualsivoglia conclusione dall'analisi diretta degli osservabili che introdurremo, occorre comprendere dopo quanto tempo il sistema raggiunge l'equilibrio conformazionale della sua struttura terziaria per poi considerare nelle medie solo i valori successivi a tale tempo di equilibratura.

Per non appesantire troppo la scrittura ci concentreremo esclusivamente sul sistema a singolo DNA della sequenza ACC contenuta nel box di volume 1.0x, verificando che le stime temporali ottenute in questo modo siano ben allineate anche nel caso di simulazioni a volumi differenti.

Poiché stiamo simulando la dinamica di un singolo DNA all'interno di un box cubico abbastanza piccolo la molecola rischia di vedere la sua immagine per colpa delle interazioni calcolate secondo le condizioni al bordo di Born Von Karmann. Per scongiurare ciò, o quantomeno per limitare l'effetto dell'interazione elettrostatica con la sua immagine periodica, è utile effettuare lo studio dei tempi di equilibratura in funzione della concentrazione salina disciolta in soluzione.

Considereremo quindi delle simulazioni effettuate all'interno di quattro box di volume 1.0x contenenti un singolo DNA di tipo ACC immerso in soluzione a concentrazioni saline differenti.

Per prima cosa andiamo a calcolare mediante il programma di analisi `g_gyrate` il raggio di girazione della molecola ("radius of gyration") definito come:

$$R_g(t) = \sqrt{\frac{\sum_i \|\mathbf{r}_i(t)\|^2 m_i}{\sum_i m_i}}. \quad (7.8)$$

Questa quantità ci permette di comprendere la compattezza di una molecola rispetto al suo centro di massa, un indicatore di quando la molecola ha raggiunto la sua struttura terziaria prevista spesso utilizzato per caratterizzare polimeri o proteine in soluzione. Dall'analisi del grafico notiamo che il sistema immerso in soluzione acquosa con concentrazione salina pari a  $10mM$  nella durata di 200ns non raggiunge nessuna configurazione di equilibrio. Al crescere della concentrazione salina notiamo che sia per  $100mM$  che per  $1000mM$  il raggio di girazione comincia a stabilizzarsi dopo circa  $100ns$  di simulazione.

In secondo luogo è utile anche valutare la radice dello scarto quadratico medio ("root mean square deviations") della struttura rispetto alla struttura di riferimento iniziale, ottenuta calcolando la somma di tutti gli spostamenti quadratici medi di ogni singolo atomo dell'acido

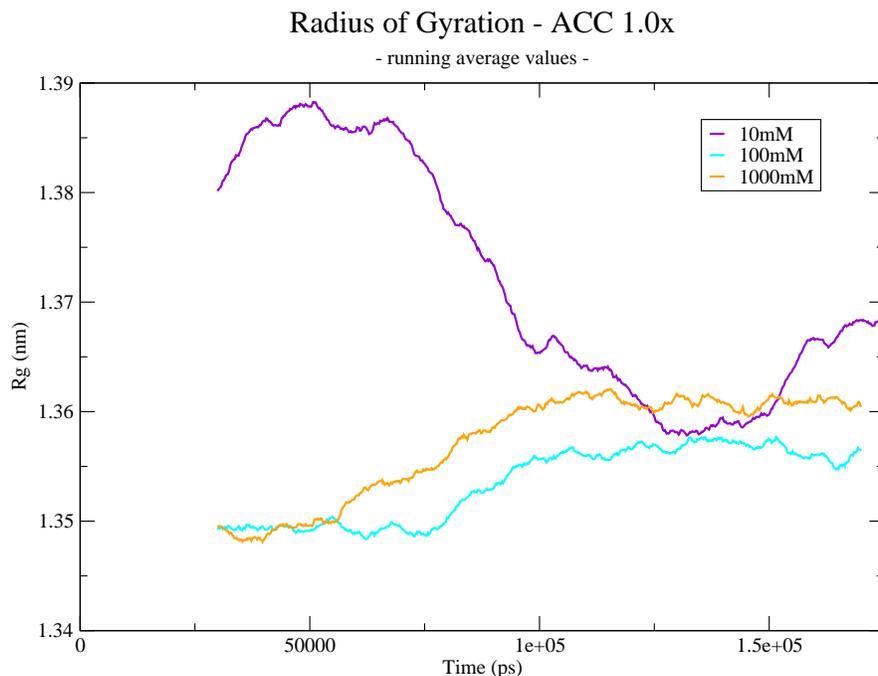


Figura 7.4: Raggio di girazione in funzione del tempo, il grafico riporta i valori delle medie mobili fatte in un intorno di  $25ns$ . Notiamo la mancata stabilizzazione della soluzione da  $10mM$ .

nucleico in funzione del tempo. Questa analisi va fatta necessariamente dopo aver eseguito un fit rototraslazionale della molecola rispetto alla posizione iniziale, per evitare di inserire nello scarto quadratico medio la distanza dovuta alla diffusione della molecola nel sistema. Tale grandezza, anch'essa funzione del tempo, è definita secondo la relazione:

$$RMSD(t_1, t_2) = \left[ \frac{1}{M} \sum_{i=1}^N m_i \|\mathbf{r}_i(t_1) - \mathbf{r}_i(t_2)\|^2 \right]^{\frac{1}{2}} \quad (7.9)$$

e calcolata attraverso il programma `g_rms` del pacchetto GROMACS. Grazie all'analisi di questa quantità possiamo stabilire effettivamente dove la forma della molecola comincia a stabilizzare in una struttura ben definita. Infatti dal grafico in figura 7.5 notiamo che la sequenza immersa in soluzione  $1M$  raggiunge il "plateau" della media della RMSD a circa  $100ns$  mentre la soluzione  $100mM$  a circa  $125ns$ . L'altra sequenza invece non stabilizza affatto ed anche esplorando tempi più lunghi fino a  $300ns$  tale quantità fluttua in modo molto ampio, non attestandosi mai attorno ad una situazione di equilibrio, chiaro indice del fatto che la schermatura elettrostatica è troppo debole.

Come ulteriore controllo analizziamo il Mean Square Displacement (MSD) usando la fun-

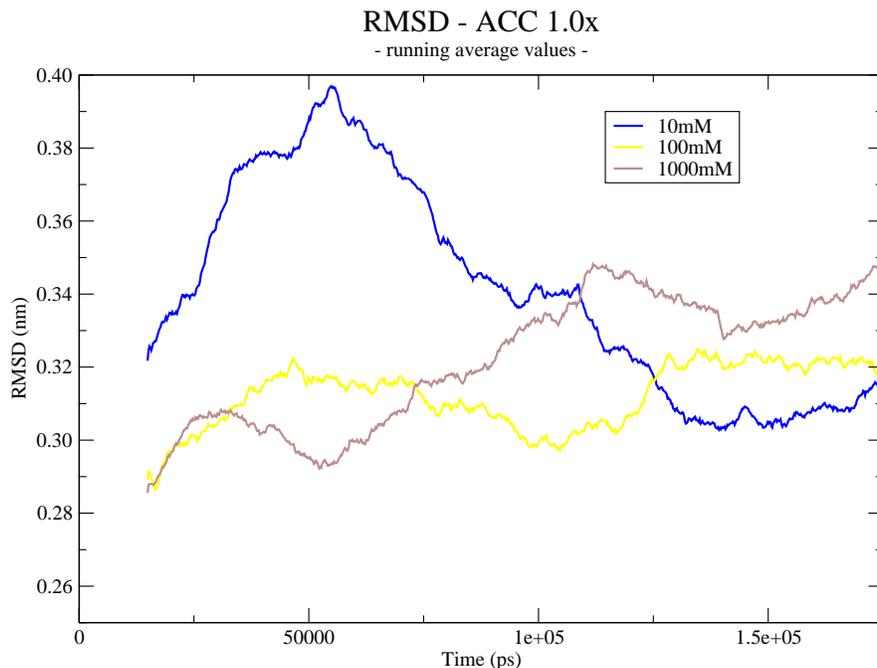


Figura 7.5: Medie delle radici degli spostamenti quadratici medi. Nel grafico non vengono riportati i valori istantanei, perché affetti da grandi fluttuazioni anche se in realtà mostrano bene il raggiungimento del “plateau” per le due. La stabilizzazione per la sequenza  $10mM$  non avviene nemmeno a tempi più lunghi.

zione `g_msd` implementata nel pacchetto GROMACS, utile per stimare i tempi entro i quali il sistema comincia a diffondere. La teoria per il calcolo del coefficiente di autodiffusione  $D_A$  porta a scrivere la relazione di Einstein:

$$\lim_{t \rightarrow \infty} \left\langle |\mathbf{r}_i(t) - \mathbf{r}_i(0)|^2 \right\rangle_{i \in DNA} = 6D_A t. \quad (7.10)$$

che ci permette di osservare che il sistema diffonde linearmente con il passare del tempo all’interno del mezzo. Per le tre sequenze, eseguendo un fit lineare dei valori mediante il programma `g_msd`, otteniamo i seguenti risultati per i coefficienti di autodiffusione:

$$D_{10mM}^{ACC} = (0.2660 \pm 0.0082) \cdot 10^{-5} \frac{cm^2}{s} \quad (7.11)$$

$$D_{100mM}^{ACC} = (0.0994 \pm 0.1766) \cdot 10^{-5} \frac{cm^2}{s} \quad (7.12)$$

$$D_{1000mM}^{ACC} = (0.1214 \pm 0.0609) \cdot 10^{-5} \frac{cm^2}{s}. \quad (7.13)$$

Perciò per quanto riguarda i tempi caratteristici con cui il sistema raggiunge una configurazione stabile, possiamo pensare di stimarli come il reciproco del coefficiente di diffusione, previa opportuna scelta delle unità di misura:

$$t \sim \frac{(1nm)^2}{D_A} \rightarrow t_{diffusione} \sim 10ns \quad (7.14)$$

ed alla luce di ciò possiamo dedurre che i tre sistemi analizzati diffondono dopo tempi dell'ordine dei  $10ns$ .

Da quanto ottenuto in queste tre simulazioni possiamo considerare  $t_{equil} = 100ns$  il tempo minimo di equilibratura necessaria per consentire all'acido nucleico di stabilizzarsi nella conformazione strutturale terziaria più probabile. Dopo questo tempo infatti il sistema ha diffuso e si è stabilizzato morfologicamente, inoltre notiamo che l'accorgimento di schermare la carica elettrostatica si rivela efficiente anche dal punto di vista computazionale per concentrazioni saline particolarmente elevate: da ora in avanti considereremo tutte simulazioni di dinamica molecolare degli acidi nucleici immersi in solvente a concentrazione salina  $1M$ .

In linea con quanto detto sinora, anche negli esperimenti reali il gruppo di Nakata [17, 16] usa concentrazioni saline estremamente alte, sempre superiori allo  $0.5M$ , che si innalzano ulteriormente durante il processo di disidratazione. Nel lavoro computazionale di Maffeo [21] invece, le simulazioni sono state effettuate in soluzioni da  $100mM$ , ma il loro lavoro era incentrato principalmente nel comprendere e quantificare le forze di “stacking end-to-end” tra due filamenti di DNA, non tanto nel comprenderne la sua struttura.

### 7.3 Convinciamoci dell'equilibratura

Anticipiamo alcuni risultati che andremo ad analizzare nel seguito, per constatare che quanto detto riguardo la tempistica di equilibratura funziona. Osservando il raggio di girazione e lo spostamento quadratico medio di un sistema composto da un solo DNA immerso in soluzione acquosa con ioni in concentrazione  $1M$ , avevamo stabilito che  $100ns$  fosse un tempo ragionevole per far aggiustare al sistema tutte le interazioni tra particelle permettendo alle osservabili di stabilizzarsi.

In seguito avevamo supposto che tutto questo funzionasse sia per box a volumi differenti e sia per box contenenti molteplici DNA. A questo punto, oltre ad aver osservato una buona stabilizzazione delle medie temporali delle osservabili da noi definite, possiamo convincerci graficamente del fatto che questa stima temporale sia ragionevole anche per un sistema a molti DNA.

Nel grafico 7.6 relativo alla distanza “end-to-end” istantanea, calcolata mediando sulle singole configurazioni istantanee di tutte le sequenze nel box, apprezziamo graficamente la stabilizzazione dei sistemi a tempi del tutto confrontabili col valore da noi ottenuto. Specialmente

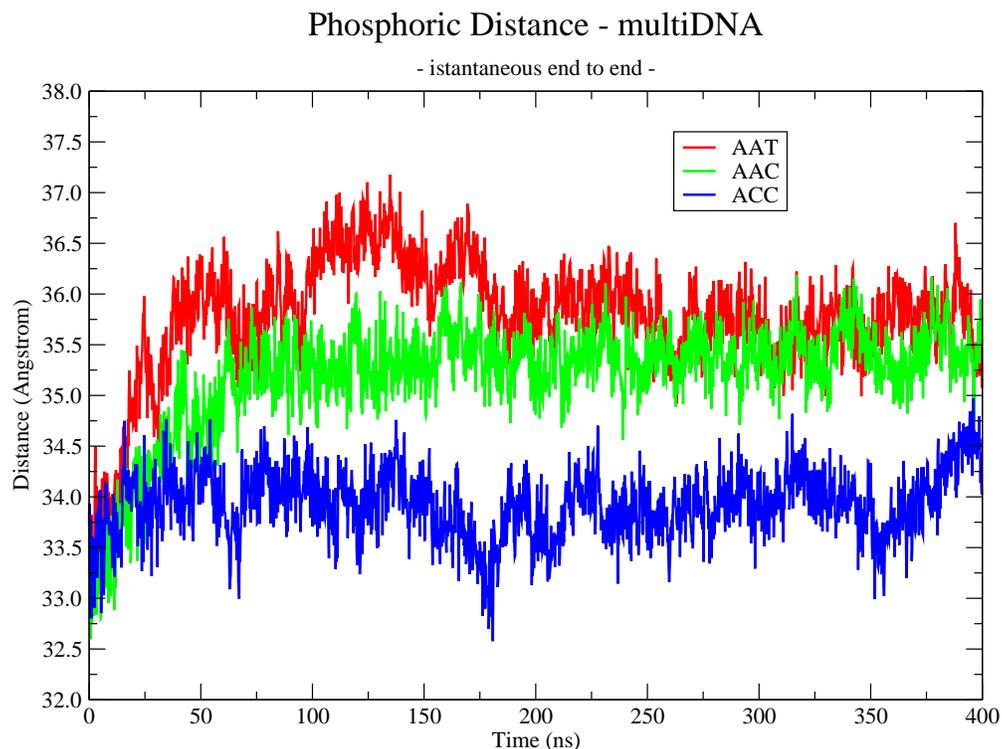


Figura 7.6: Valori istantanei delle elongazioni delle tre sequenze considerate. Notiamo come  $100\text{ns}$  rappresentino anche in questo caso una buona stima dei tempi di equilibratura per i tre sistemi.

per la sequenza AAT, risulta evidente come i doppi filamenti partendo da una configurazione iniziale lievemente più compressa si allungano fino a raggiungere la lunghezza che gli consente la stabilizzazione.

Inoltre per le sequenze AAT ed ACC osserviamo delle variazioni repentine della distanza ad alcuni istanti, questo probabilmente è dovuto a fluttuazioni in cui il sistema passa da alcune configurazioni stabili ad altre.

## Capitolo 8

# Definizioni delle osservabili

Stimolati da quanto visto per la densità a cui i sistemi stabilizzano, con particolare riferimento alla simulazione dei sistemi contenenti molti DNA, passiamo all'introduzione delle osservabili che andremo ad esaminare durante nel capitolo relativo all'analisi dati.

A causa della geometria elicoidale diventa complesso capire in che modo il DNA si piega. A causa del fatto che dopo circa 11 basi azotate l'acido nucleico si è completamente avvitato su se stesso, è ragionevole pensare che gran parte dell'effetto della piegatura sia localizzato lungo il solco maggiore. A causa di questo fatto, nonostante la sequenza sia palindroma e quindi speculare rispetto al suo centro geometrico come disposizione delle basi, non lo è rigorosamente dal punto di vista geometrico.

Ora poiché il modello che introdurremo nel seguito sarà a doppio cilindro piegato, trascureremo questa asimmetria ed andremo a suddividere il DNA in due parti composte da 6 coppie di basi azotate e ciascuna di esse verrà considerata come approssimabile ad un cilindro.

Ispirati dal lavoro di Dickerson [44] e dalla definizione dei parametri FREEHELIX in primo luogo andiamo a considerare due grandezze relative a due coppie di basi successive:

1. Angolo tra la normale a due coppie di basi successive;
2. Distanza tra i centri geometrici di due coppie di basi successive.

Dal punto di vista geometrico andremo a scegliere, per ciascuna coppia di basi, una terna di atomi che definiranno la base del sistema di riferimento. Gli atomi che definiscono la base sono scelti basandoci sul modello di [44] e sono quelli denominati C1', N6 ed N4. Su ciascuna di queste terne di atomi faremo passare un piano che rappresenta il piano della coppia di basi azotate e ne calcoleremo il baricentro. A questo punto calcoleremo gli angoli diedrali tra piani successivi definiti come gli angoli formati tra le normali alla coppia di piani considerati. Questo ci fornirà l'angolo di inclinazione totale relativo alle due coppie di basi, denominato VALL nel testo di riferimento, determinato dalla combinazione degli effetti di "tilting" (VTIL)

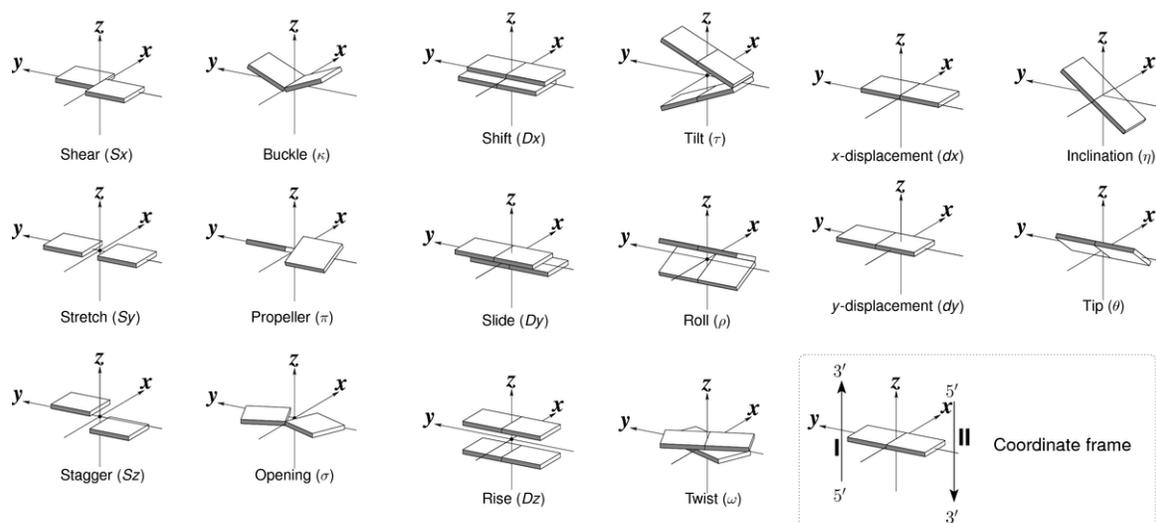


Figura 8.1: Tutti i possibili parametri FREEHELIX standard definiti nel lavoro di Dikerson [44]. Il gruppo dei sei parametri centrali sono quelli presi in esame per valutare la morfologia delle sequenze da noi considerate.

e “rolling” (VROL) tra le coppie di piani come mostrato in figura 8.1. Approssimativamente per piccoli angoli vale la relazione:

$$VALL^2 = VTIL^2 + VROL^2. \quad (8.1)$$

Notiamo inoltre che così facendo stiamo trascurando il contributo di “twisting” poiché non siamo interessati ad esaminare il contributo di torsione del sistema lungo l’asse.

Il calcolo della distanza geometrica tra i centri geometrici della terna di atomi considerati invece è legato alla somma degli effetti di “shifting” (VSHF), “sliding” (VSLI) e “rising” (VRIS) e verrà denominato VDIST anche se non presente nella definizione delle osservabili di Dickerson. Anche in questo caso il contributo di VDIST si lega agli altri parametri attraverso la relazione, stavolta non approssimata:

$$VDIST^2 = VSHF^2 + VSLI^2 + VRIS^2. \quad (8.2)$$

I parametri appena introdotti serviranno come indicatori del fatto che è presente un bending differente nei sistemi composti da diverse sequenze di DNA, ma non ci forniranno un angolo complessivo valido per modellizzare il sistema attraverso la coppia di cilindri piegati. A questo scopo introduciamo due nuove osservabili angolari che ci permetteranno di quantificare il fenomeno del “bending”, ricordando valida l’approssimazione del DNA in due sotto-segmenti uguali:

1. Angolo di “bending” assiale;
2. Angolo “bending” dei baricentri dei gruppi fosforici;
3. Distanza “end-to-end” tra baricentri di gruppi fosforici.

Entriamo più nel dettaglio per definire queste nuove grandezze appena introdotte.

## 8.1 Il “bending” assiale

La prima delle osservabili introdotte, il “bending” assiale, è modellizzata in modo del tutto analogo alla definizione dell’asse principale (“viewing-axis”) descritto in [44]. Nel modello scelto vengono presi gli atomi C1’ di ogni base ed N1 o N9, a seconda se la base azotata sia timina o citosina (N1) oppure adenina o guanina (N9). A differenza del modello di Dickerson che considera i vettori congiungenti gli atomi C1’-C1’ ed N-N di basi successive, in questa sede sfrutteremo i vettori che congiungono i baricentri degli atomi C1’ ed N di ciascuna coppia di basi. Calcoleremo perciò i baricentri degli atomi C1’ tra ciascuna base azotata e la controparte con cui essa è accoppiata, ripetendo la stessa operazione per gli N. In questo modo avremo definito 24 baricentri, 12 per ciascuna delle due metà in cui abbiamo suddiviso l’acido nucleico attraverso i quali calcoleremo i 10 vettori normalizzati che li congiungono, 5 per la metà superiore ed altrettanti per la inferiore.

A causa della presenza della stessa coppia di basi alle estremità di ogni sequenza simulata, dovremo tener conto di un particolare effetto dovuto alla debolezza del “base pairing” della coppia A-T mediante doppio legame ad idrogeno: il “fraying” [31]. Rispetto alla coppia G-C che è stabilizzata da tre legami ad idrogeno, la connessione A-T è più debole e ad energia ambiente è più facilmente soggetta ad un fenomeno di sfilacciamento. La rottura della coppia determina una sostanziale modifica nella struttura dell’acido nucleico, fatto che si è rivelato estremamente dannoso durante l’analisi dati. Tale fenomeno ovviamente non è presente durante il processo di “self-assembly” che porta alla formazione della fase LC, poiché l’interazione di “stacking” stabilizza la coppia favorendone l’impilamento e si manifesta principalmente se la coppia A-T si trova in posizione estrema,

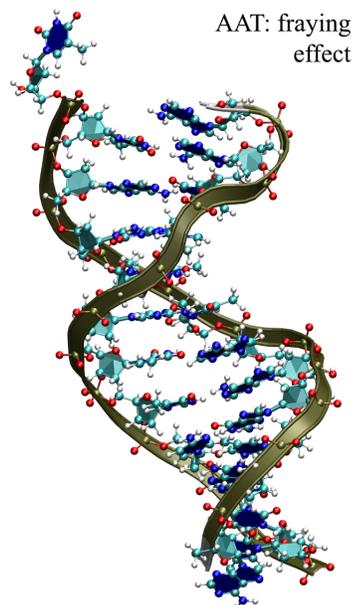


Figura 8.2: Fenomeno di “fraying” dell’ultima coppia di basi della sequenza AAT. La struttura viene deformata dallo sfilacciamento.



gsl multifit linear workspace \*work)

che esegue il fit di dati usando un modello lineare generico della forma:

$$\mathbf{z} = M\mathbf{c} \quad (8.3)$$

dove  $\mathbf{z}$  è il vettore di  $n$  osservazioni,  $M$  è una matrice  $n \times p$  che contiene le variabili di predizione e  $\mathbf{c}$  contiene i  $p$  elementi che meglio approssimano la funzione della quale vogliamo eseguire il fit lineare.

Nel nostro caso particolare vogliamo approssimare i dati con una funzione planare del tipo:

$$ax + by + cz + d = 0 \quad (8.4)$$

ora fissando il parametro  $c = -1$ , scriviamo la funzione della quale eseguire il fit:

$$f(x, y) = z = ax + by + d \quad (8.5)$$

che ci consente di implementare il modello lineare planare nella forma:

$$z = \begin{pmatrix} x & y & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ d \end{pmatrix} \quad (8.6)$$

dove la matrice  $M = \begin{pmatrix} x & y & 1 \end{pmatrix}$  contiene i valori delle coordinate  $x$  ed  $y$  di ciascun vettore che congiunge i baricentri dei  $C1'$  (e successivamente che congiunge i baricentri degli N),  $z$  è la terza componente e  $\mathbf{c} = \begin{pmatrix} a \\ b \\ d \end{pmatrix}$  contiene i valori che meglio approssimano la funzione  $f(x, y)$ .

Il coefficiente  $d$  è inutile ai fini del calcolo, poiché il vettore  $\mathbf{v}_{\perp}^{plane}$  normalizzato e perpendicolare a ciascun piano sarà dato da:

$$\mathbf{v}_{\perp}^{plane} = \frac{\mathbf{w}^{plane}}{|\mathbf{w}^{plane}|} = \frac{1}{|\mathbf{w}^{plane}|} \begin{pmatrix} a \\ b \\ 1 \end{pmatrix}. \quad (8.7)$$

C'è da notare che con questa forma non è possibile fittare un piano il cui vettore perpendicolare è esso stesso perpendicolare all'asse  $z$ , poiché non sarebbe possibile calcolare la funzione  $f(x, y)$ . Questa eventualità viene sostanzialmente eliminata mediante una continua trasformazione di coordinate che ruotano il DNA mantenendo il "viewing-axis" globale sostanzialmente parallelo all'asse  $z$ .

Di tali piani calcoliamo i vettori normali ad essi associati per poi sommarli a due a due per ciascuna metà normalizzando nuovamente. In questo modo abbiamo ottenuto due vettori normalizzati, uno che rappresenta la direzione che meglio approssima l'asse del sotto-segno inferiore, l'altro che meglio approssima quello superiore. Tali vettori rappresentano perciò gli assi di ciascuna delle due metà in cui avevamo supposto poter dividere il DNA, ed il calcolo dell'angolo formato tra questi assi porterà ad una prima stima del grado di "bending" assiale.

Purtroppo però questa può essere solamente una stima indicativa della piegatura reale della struttura, poiché vengono escluse dal calcolo le coppie A-T alle estremità dell'acido nucleico a causa del fenomeno di "fraying": a nematizzazione avvenuta le basi risulteranno accoppiate poiché le forze di "stacking" saranno sufficienti a compensare tale effetto ed il "bending" coinvolgerà tutta la struttura molecolare, non solo una parte come analizzato dall'osservabile appena introdotta.

## 8.2 Baricentri dei gruppi fosforici

Per poter spingere l'analisi fin quasi a giungere alle coppie di basi estremali ma senza incorrere nel fenomeno del "fraying", possiamo pensare di ideare un nuovo osservabile, stavolta concentrando la nostra attenzione sul backbone fosforico degli acidi nucleici.

Decidiamo per prima cosa di costruire uno scheletro interno del polimero, calcolando i baricentri di ciascuna coppia di gruppi fosforici congiungenti due coppie di nucleobasi. A causa della geometria a due solchi principali lo scheletro interno dei baricentri dei gruppi fosforici si disporrà su una geometria elicoidale, avvitandosi anch'esso attorno ad un asse principale.

Nel caso delle nostre sequenze palindrome la scelta di questa osservabile si rivela particolarmente efficiente: essendo formati da 12 coppie di basi azotate, gli acidi nucleici presi in esame presentano  $11 \cdot 2 = 22$  gruppi fosforici dei quali la 6° coppia si trova esattamente a metà dell'altezza della molecola. Non solo. Sappiamo che il B-DNA compie un avvitamento completo dopo circa 11 basi azotate, perciò le coppie fosforiche iniziali e finali si trovano esattamente allineate l'una sopra l'altra, come pure i baricentri.

Da queste caratteristiche risulta abbastanza naturale definire il "bending" della molecola legandolo all'angolo che congiunge i baricentri dei gruppi fosforici estremali con il baricentro del gruppo centrale. Così facendo stiamo osservando di quanto lo scheletro interno della molecola si chiude sul solco maggiore (in gran parte l'effetto di piegatura si riscontrerà esattamente in quella zona) e rappresenta una buona stima di quello che accade alla struttura della molecola. La distanza tra il primo e l'ultimo baricentro, invece, potrà essere in buona approssimazione legata alla lunghezza "end-to-end" dell'acido nucleico, ovvero la distanza effettiva tra le basi terminali, che definisce un quindi un terzo osservabile sperimentale.

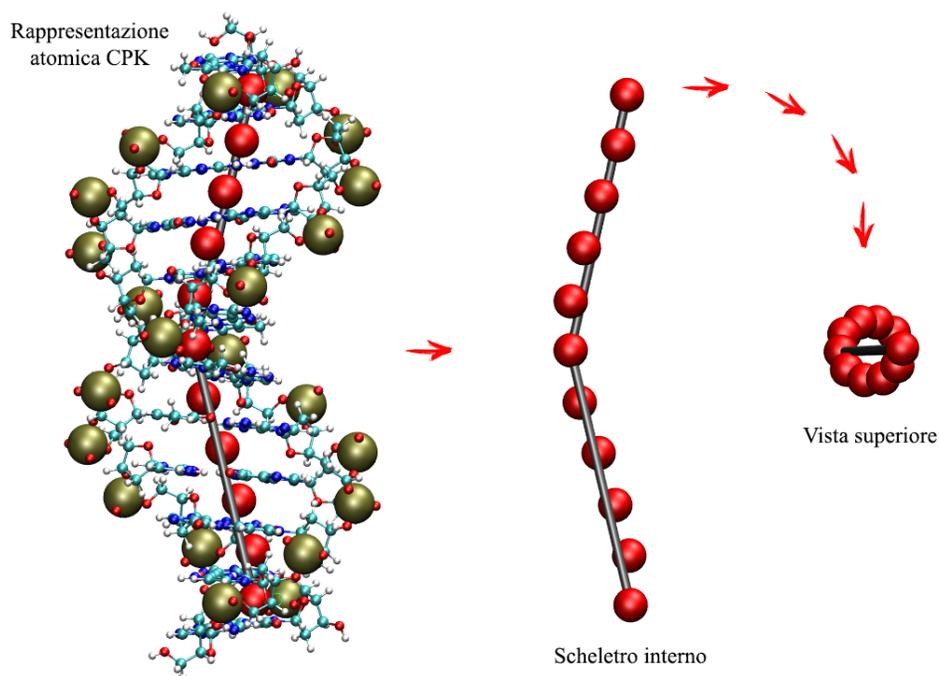


Figura 8.4: Ancora una volta la figura mostra una sequenza AAT. I baricentri delle coppie di gruppi fosforici sono le grandi sfere rosse, mentre l'angolo al centro appena definito (A2C) è l'angolo tra i vettori rappresentati in verde.

Poiché abbiamo detto che questa piegatura è legata al “bending” dell'intera struttura del DNA, possiamo pensare di creare un modello che lo leghi direttamente all'angolo tra gli assi dei cilindri che useremo per modellizzare la struttura a doppia elica.

Per prima cosa chiamiamo il baricentro fosforico centrale  $A$  e quelli agli estremi  $B$  e  $C$ . Successivamente consideriamo il segmento  $\overline{BC} = d$  che abbiamo detto essere la distanza “end-to-end” ed  $H$  la proiezione perpendicolare ad esso del punto  $A$ . L'angolo tra i baricentri agli estremi con quello centrale viene chiamato  $B\hat{A}C = \theta$  mentre l'angolo effettivo di piegatura sarà  $B\hat{H}C = \Theta$ . A questo punto abbiamo tutti i valori geometrici iniziali  $\overline{AH}_i = h_i$ ,  $\overline{BC}_i = d_i$ ,  $\theta_i$  e  $\Theta_i$  determinati dalla struttura cristallografica di partenza ed a simulazioni ultimate saremo in grado di stimare i valori medi di  $\theta_M$  ed  $\overline{BC}_M = d_M$ . Occorre perciò calcolare l'angolo di piegatura equivalente medio  $\Theta_M$ .

Consideriamo in prima approssimazione fissata la lunghezza del segmento  $\overline{AH} = h$ , come pure il fatto che esso sia bisettrice degli angoli  $\theta$  e  $\Theta$  per ogni tempo: sappiamo che questo non è vero a causa della presenza delle diverse ampiezze dei solchi maggiore e minore, che implica una sostanziale modifica a tale geometria poiché il centro di piegatura potrebbe non essere

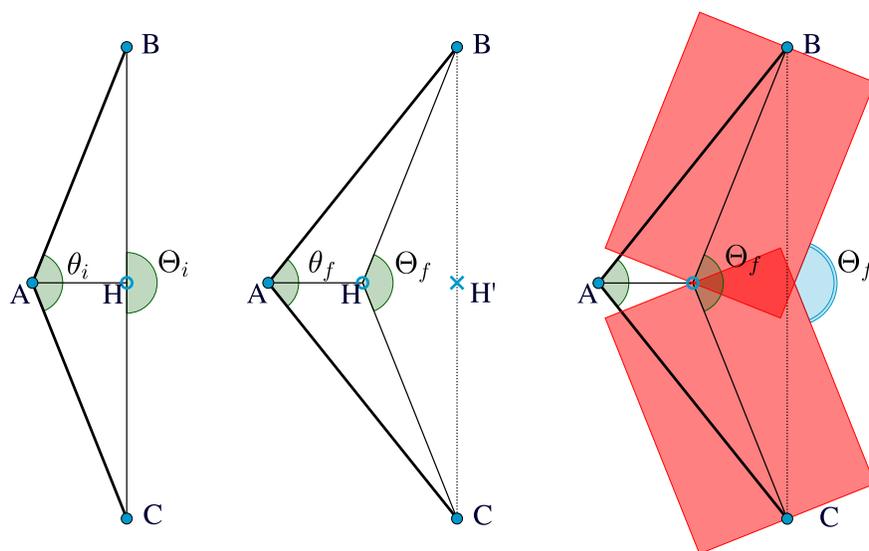


Figura 8.5: Modellino che illustra concettualmente come legare il bending fosforico al bending assiale del segmento di DNA per poi ottenere quindi la schematizzazione a doppio cilindro piegato.

localizzato nel baricentro del gruppo fosforico centrale. Il punto medio del segmento  $CB$  dopo il calcolo dei valori medi viene chiamato  $H'$  di modo tale che  $\overline{AH'} = h'$ .

Per prima cosa sappiamo che l'angolo tra i baricentri può essere calcolato come:

$$\theta = \arccos \left( \frac{\mathbf{v}_{P1} \cdot \mathbf{v}_{P2}}{|\mathbf{v}_{P1}| |\mathbf{v}_{P2}|} \right) \quad (8.8)$$

dove  $\mathbf{v}_{P1} = \overrightarrow{AB}$  è il vettore che congiunge il baricentro del gruppo centrale al gruppo fosforico superiore ed analogamente  $\mathbf{v}_{P2} = \overrightarrow{AC}$  per quello inferiore. Con alcuni passaggi di trigonometria elementare otteniamo:

$$i \sin \left( \frac{\theta}{2} \right) = \frac{d}{2} \quad \rightarrow \quad i = \frac{d}{2 \sin \left( \frac{\theta}{2} \right)} \quad (8.9)$$

che usiamo per calcolare  $h$  ed  $h'$ :

$$h = i \cos \left( \frac{\theta_i}{2} \right) = \frac{d_i}{2} \cot \left( \frac{\theta_i}{2} \right); \quad h' = \frac{d_M}{2} \cot \left( \frac{\theta_M}{2} \right). \quad (8.10)$$

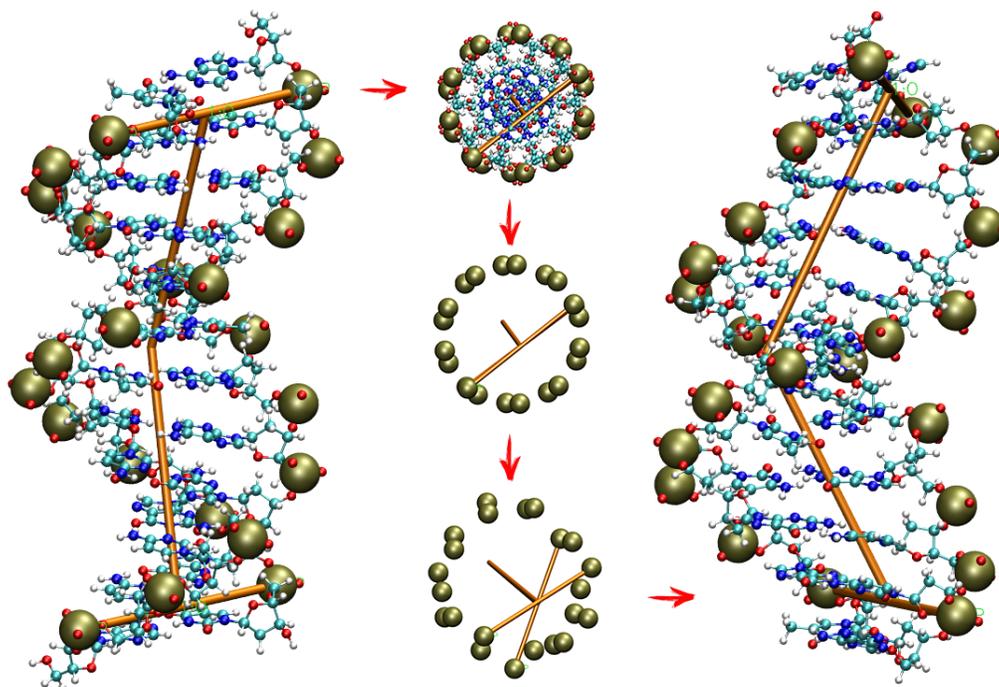


Figura 8.6: A titolo di esempio è riportata la sequenza AATGAATTCATT. Gli assi che compaiono nell'immagine sono le distanze che vengono prese in esame dal software di analisi mentre le sfere dorate rappresentano le posizioni degli atomi di fosforo. Nella visuale dall'alto è possibile osservare l'angolo di "twisting" completo della molecola attorno all'asse principale.

Ora consideriamo il nuovo segmento  $\overline{HH'}$ :

$$\overline{HH'} = \overline{AH'} - \overline{AH} = h' - h = \frac{1}{2} \left[ d_M \cot \left( \frac{\theta_M}{2} \right) - d_i \cot \left( \frac{\theta_i}{2} \right) \right] \quad (8.11)$$

che leghiamo all'angolo di "bending" equivalente:

$$\cot \left( \frac{\Theta_M}{2} \right) = \frac{\overline{HH'}}{\overline{AB_M/2}} = 2 \frac{h' - h}{d_M} \quad (8.12)$$

e che mi permette di scrivere:

$$\cot \left( \frac{\Theta_M}{2} \right) = \cot \left( \frac{\theta_M}{2} \right) - \frac{d_i}{d_M} \cot \left( \frac{\theta_i}{2} \right). \quad (8.13)$$

A questo punto basta semplicemente calcolare la funzione inversa per esplicitare l'angolo

di bending equivalente mediante la relazione:

$$\Theta_M = 2 \cot^{-1} \left\{ \cot \left( \frac{\theta_M}{2} \right) - \frac{d_i}{d_M} \cot \left( \frac{\theta_i}{2} \right) \right\} \quad (8.14)$$

come funzione dei soli parametri iniziali  $d_i$ ,  $\theta_i$  e medi  $d_M$ ,  $\theta_M$ .

Questa osservabile si rivelerà particolarmente utile poiché per come è costruita è affetta da fluttuazioni più contenute rispetto alla precedente osservabile e ci consente di apprezzare ancora una volta la differenza di bending tra le diverse sequenze. Il programma che analizza tale grado di piegatura è stato scritto in C++ come nel caso precedente e permette di calcolare altre grandezze utili, come per esempio il “twisting” completo della molecola e le inclinazioni della prima ed ultima coppia di gruppi fosforici, definendo un vettore che congiunge un gruppo fosforico al corrispondente gruppo posto sull’altro filamento.

## Capitolo 9

# Analisi delle simulazioni single-DNA

### 9.1 Effetto di taglia del sistema

Simulare il comportamento di un sistema dinamico con un gran numero di costituenti è una sfida computazionale piuttosto ardua. Con i mezzi a disposizione e con le attuali potenze di calcolo una simulazione con un numero statisticamente rilevante di DNA immersi in soluzione è pressoché impossibile da effettuare in tempi finiti.

Poiché siamo interessati ad ottenere delle proprietà strutturali che caratterizzino la singola sequenza, possiamo pensare in primo luogo di studiare il comportamento di un sistema composto da un singolo segmento di DNA immerso in soluzione acquosa a concentrazione 1M di *NaCl* contenuto in un box cubico. Il problema principale di una simulazione di questo tipo è che il polimero potrebbe, a causa delle condizioni periodiche al contorno, interagire con la sua immagine creando degli artefatti nella simulazione che influenzano la dinamica del sistema.

Decidiamo quindi di agire in questo modo: consideriamo al solito le  $t$  sequenze di DNA (AAT, AAC, ACC) e simuliamole ciascuna in box cubici a diverso volume (1.0x, 2.0x, 3.0x), come definito nel capitolo precedente. C'è da considerare che le concentrazioni dell'acido nucleico a cui lavoriamo sono estremamente più basse di quelle riscontrate per la transizione di fase da Nakata e Zanchetta [16], che stimano per le concentrazioni valori attorno a  $c_{DNA}^{nem} > 500mg/ml$ . Sappiamo che il peso molecolare di un singola base azotata è pari circa a  $303.7u$ , perciò la massa totale approssimata dei doppi filamenti presi in esame sarà:

$$m_{DNA} = 2 (\#basi) \cdot 303.7u = 7288.8u = 1.210 \cdot 10^{-17}mg \quad (9.1)$$

poiché l'unità di massa atomica vale circa  $1u \approx 1,660538921 \cdot 10^{-21}mg$ . A questo punto è facile calcolarsi la concentrazione del DNA in funzione del box in cui è immerso mediante la

formula  $c_{DNA} = \frac{m_{DNA}}{V_{box}}$ :

$$c_{DNA}^{1.0x} = 53.9 \frac{mg}{ml}; \quad c_{DNA}^{2.0x} = 27.0 \frac{mg}{ml}; \quad c_{DNA}^{3.0x} = 18.0 \frac{mg}{ml}. \quad (9.2)$$

Come vediamo queste concentrazioni sono un ordine di grandezza più piccole rispetto alle concentrazioni critiche a cui si riscontra la transizione di fase nematica. Purtroppo però non è possibile aumentare ulteriormente la concentrazione critica per alcuni motivi:

1. La struttura allungata della molecola di DNA impedisce di fare una simulazione all'interno di un box a forma di parallelepipedo poiché durante l'evoluzione temporale essa potrebbe ruotare disallineandosi rispetto al box, interagendo direttamente con se stessa nonostante opportune scelte di cut-off. Il box deve essere quindi cubico.
2. Diminuendo la lunghezza del box cubico potremmo incorrere in un altro problema legato alle condizioni periodiche, in questo caso la coppia di basi inferiore "vede" troppo vicina la coppia di basi superiore della molecola ottenuta dalla "periodic boundary condition" e viceversa. Anche in questo caso quindi è necessario scegliere la dimensione del box cubico in modo tale che il doppio della distanza delle basi estremali dalle facce del box ad esse parallele sia superiore delle dimensione del cut-off che abbiamo settato pari a  $d_{cut-off} = 1nm$ .

Vedremo ora come tutto questo influenzi la dinamica dei sistemi. Siccome non abbiamo molti dati a disposizione in questo caso non è interessante fare un'analisi statistica delle osservabili. Supponendo i sistemi equilibrati dopo  $100ns$  graficheremo nelle figure il valor medio ottenuto dal calcolo degli angoli assiali e fosforici degli osservabili definiti nel capitolo precedente.

Usando come ulteriore criterio di equilibratura delle strutture anche la distanza "end-to-end" delle basi estremali della molecola notiamo che in  $300ns$  di simulazione riescono ad equilibrare solamente le sequenze contenute all'interno del volume 1.0x. Per le altre (salvo la AAC nel volume 2.0x) i tempi di equilibratura sono più lunghi per colpa del fatto che il sistema è composto da un maggior numero di atomi.

Tenendo presente questo fatto andiamo ad osservare, al variare della dimensione del box, su quale valore medio si attestano le osservabili. Notiamo che nel volume 1.0x la correlazione è ben riprodotta per entrambe le osservabili: La sequenza AAT, che nematizza a concentrazioni inferiori, è quella che risulta meno piegata e più allungata delle altre, la AAC ha come previsto un comportamento intermedio mentre la ACC è quella che risulta effettivamente più piegata e compressa.

Le cose cominciano a cambiare nel caso dei run nel volume 2.0x. La ACC come previsto risulta ancora la più piegata mentre c'è una leggera inversione di tendenza tra le altre due riscontrata in entrambi gli osservabili angolari. Probabilmente questo è dovuto anche al fatto che i sistemi ancora non hanno pienamente i stabilizzato.

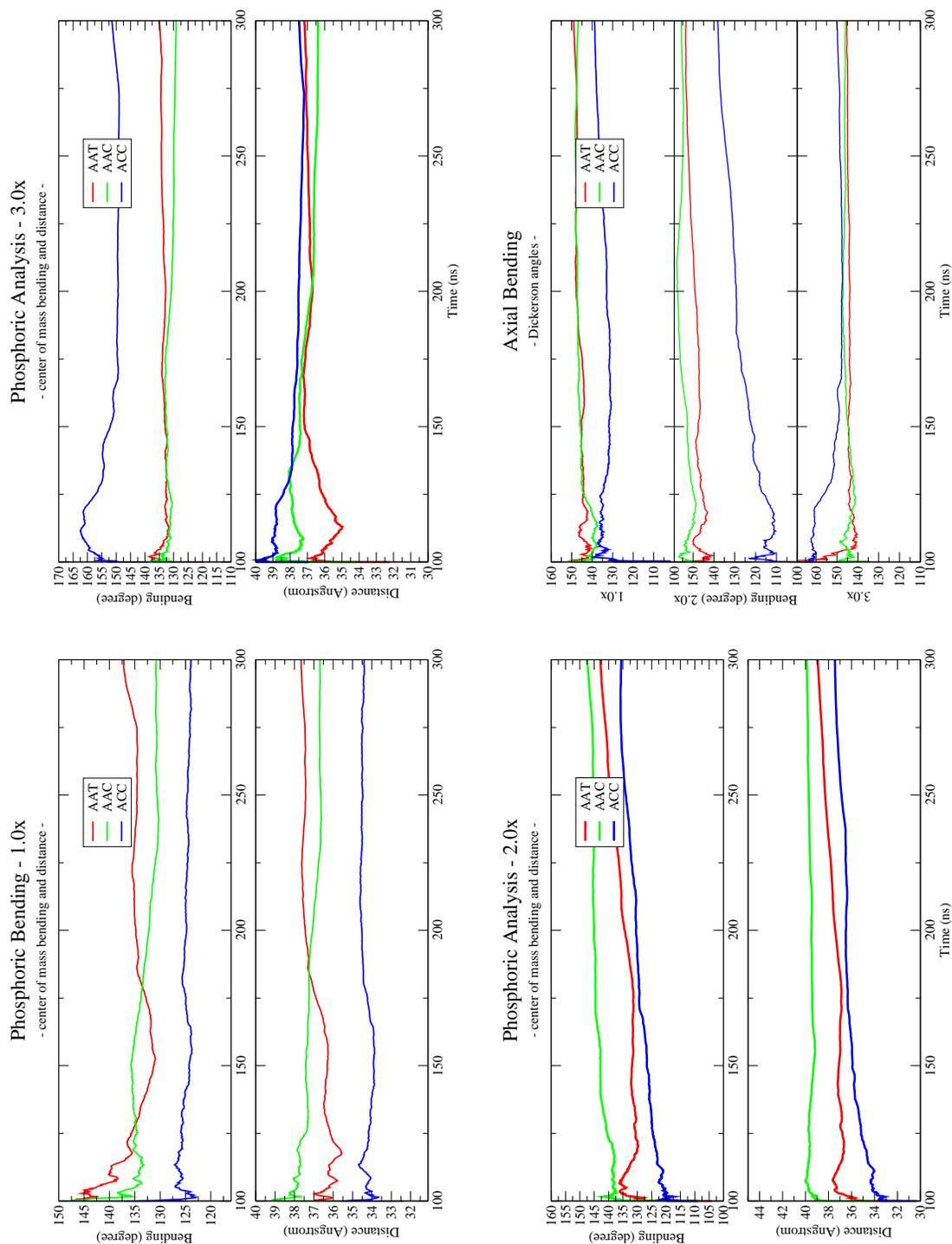


Figura 9.1: Grafici di tutti gli osservabili relativi alla misura del bending degli acidi nucleici al variare della dimensione del box di simulazione.

Nel caso estremo a bassissima concentrazione di DNA nel volume 3.0x riscontriamo una totale inversione di tendenza. In questo caso la sequenza più rettilinea risulta essere la ACC mentre le altre sono più piegate. Sembra quindi sconveniente e poco realistico simulare un sistema in queste condizioni poiché il fenomeno del bending potrebbe essere legato anche a contributi legati all'interazione con altri DNA immersi in soluzione, ovviamente annullati dal fatto che ora la molecola risente molto meno dell'interazione con la sua immagine.

I valori medi per gli angoli e per le distanze “end-to-end” sono riportati di seguito:

$$\theta_{AAT}^{1.0x} = (138.05 \pm 0.33)^\circ \quad \theta_{AAC}^{1.0x} = (132.76 \pm 0.25)^\circ \quad \theta_{ACC}^{1.0x} = (124.79 \pm 0.30)^\circ \quad (9.3)$$

$$\theta_{AAT}^{2.0x} = (138.85 \pm 0.35)^\circ \quad \theta_{AAC}^{2.0x} = (142.82 \pm 0.33)^\circ \quad \theta_{ACC}^{2.0x} = (132.81 \pm 0.31)^\circ \quad (9.4)$$

$$\theta_{AAT}^{3.0x} = (134.59 \pm 0.26)^\circ \quad \theta_{AAC}^{3.0x} = (129.04 \pm 0.24)^\circ \quad \theta_{ACC}^{3.0x} = (152.57 \pm 0.34)^\circ \quad (9.5)$$

$$d_{AAT}^{1.0x} = (37.635 \pm 0.056) \text{ \AA} \quad d_{AAC}^{1.0x} = (36.984 \pm 0.0478) \text{ \AA} \quad d_{ACC}^{1.0x} = (33.980 \pm 0.055) \text{ \AA} \quad (9.6)$$

$$d_{AAT}^{2.0x} = (38.023 \pm 0.067) \text{ \AA} \quad d_{AAC}^{2.0x} = (38.804 \pm 0.061) \text{ \AA} \quad d_{ACC}^{2.0x} = (36.787 \pm 0.061) \text{ \AA} \quad (9.7)$$

$$d_{AAT}^{3.0x} = (37.256 \pm 0.051) \text{ \AA} \quad d_{AAC}^{3.0x} = (35.803 \pm 0.053) \text{ \AA} \quad d_{ACC}^{3.0x} = (37.639 \pm 0.047) \text{ \AA} \quad (9.8)$$

$$\Theta_{AAT}^{1.0x} = (149.19 \pm 0.45)^\circ \quad \Theta_{AAC}^{1.0x} = (146.83 \pm 0.42)^\circ \quad \Theta_{ACC}^{1.0x} = (137.77 \pm 0.56)^\circ \quad (9.9)$$

$$\Theta_{AAT}^{2.0x} = (149.29 \pm 0.49)^\circ \quad \Theta_{AAC}^{2.0x} = (151.60 \pm 0.44)^\circ \quad \Theta_{ACC}^{2.0x} = (138.40 \pm 0.57)^\circ \quad (9.10)$$

$$\Theta_{AAT}^{3.0x} = (146.56 \pm 0.47)^\circ \quad \Theta_{AAC}^{3.0x} = (144.59 \pm 0.51)^\circ \quad \Theta_{ACC}^{3.0x} = (151.35 \pm 0.43)^\circ \quad (9.11)$$

Abbiamo perciò ottenuto un'indicazione del fatto che le strutture a doppia elica si aggiustano a formare una struttura piegata a determinati angoli caratteristici. Questo fenomeno sembra dovuto anche all'interazione non solo tra le sue stesse costituenti ma anche tra altre molecole che, ricordiamo, nei box a singolo DNA, corrispondono alle sue immagini. Sicuramente tutti i sistemi si sarebbero dovuti analizzare a tempi molto più lunghi, specialmente al crescere delle dimensioni dei box, per consentire a tutte le osservabili di raggiungere una struttura di equilibrio. Anche il fatto di lavorare a concentrazioni molto lontane da quelle a cui si registra la nematizzazione potrebbe aver influenzato le simulazioni come pure la possibilità della presenza di diverse configurazioni di “bending” preferite da ciascuna simulazione. Occorre dunque analizzare un sistema più realistico.

## Capitolo 10

# Analisi delle simulazioni multi-DNA

Abbiamo visto che le simulazioni a singolo DNA non forniscono grandi indicazioni sul fenomeno della piegatura e nella simulazione entravano in gioco molti parametri che rendevano il sistema poco realistico e quindi non confrontabile con un sistema reale.

Decidiamo perciò di effettuare delle simulazioni in modo che la possibilità di interazione di una molecola con se stessa sia praticamente ridotta a zero, considerando l'evoluzione a dinamica molecolare di 18 acidi nucleici immersi in soluzione acquosa a concentrazione ionica  $1M$ , considerando valide le affermazioni fatte sinora sulla velocità di equilibratura.

A causa della presenza di molti filamenti di DNA, la concentrazione degli acidi nucleici in soluzione in questo caso è molto più alta del caso precedente:

$$c_{18DNA}^{box} \simeq 324.4 \frac{mg}{ml} \quad (10.1)$$

che è un valore prossimo a quello della fase nematica dei sistemi costituiti da AAT ( $c_{AAT}^{nem} = 500 \frac{mg}{ml}$ ).

Anche stavolta le interazioni tra altre doppie eliche saranno presenti, però occorreranno sempre tra molecole differenti e la scelta della geometria  $3 \times 3 \times 2$  impedisce ad ogni molecola di vedere se stessa. Inoltre a causa del gran numero di sequenze a disposizione sarà possibile effettuare un'analisi statistica su una gran mole di dati, che ci consentirà di ottenere delle stime per il grado di bending di ciascuna sequenza.

Stavolta eseguiremo nel dettaglio l'analisi di tutte le osservabili che abbiamo a disposizione argomentando punto per punto i risultati ottenuti.

### 10.1 Angolazioni e distanza tra Basi

Per prima cosa andiamo ad esaminare l'angolazione tra i piani perpendicolari alle coppie di basi azotate e la distanza relativa ai baricentri degli stessi.

Innanzitutto, poiché tutte le sequenze sono palindrome ed abbiamo supposto la molecola simmetrica rispetto al suo centro, per non appesantire troppo l'analisi consideriamo gli angoli tra la prima e la seconda coppia nello stesso conteggio di quelli della undicesima e dodicesima. Stiamo praticamente supponendo che la simmetria della molecola si distribuisca anche in una simmetria di angolazione tra le basi:

$$\phi_{1-2} = \phi_{11-12} \quad \phi_{2-3} = \phi_{10-11} \quad \phi_{3-4} = \phi_{9-10} \quad (10.2)$$

$$\phi_{4-5} = \phi_{8-9} \quad \phi_{5-6} = \phi_{7-8} \quad \phi_{6-7} = \phi_{7-6} \quad (10.3)$$

dove l'ultima relazione per la  $\phi_{6-7}$  è ridondante poiché stiamo considerando la stessa coppia di piani.

A questo punto possiamo riportare la distribuzione di ciascuno dei sei angoli per ciascuna delle tre sequenze. I primi tre grafici in figura 10.1 riportano la distribuzione degli angoli per ogni coppia di basi a seconda del tipo di sequenza considerato, mentre dai sei grafici di figura 10.2, che riportano lo stesso angolo per le sequenze, possiamo fare alcune considerazioni importanti:

1. Analizzando i dati riguardanti l'angolo  $\phi_{1-2}$  ci accorgiamo della massiccia presenza del fenomeno del "fraying" soprattutto per le sequenze AAT ed AAC. Le prime tre coppie del doppio filamento AAT sono legate tra loro dalla presenza di soli 6 legami ad idrogeno mentre sono 7 i legami nella AAC ed 8 nella ACC. Il diverso numero di tali legami a disposizione spiega la maggior predisposizione della sequenza più rettilinea, la AAT, a subire tale fenomeno di distaccamento delle basi terminali;

$$\phi_{AAT}^{1-2} = (21.16 \pm 20.73)^\circ \quad \phi_{AAC} = (18.18 \pm 16.24)^\circ \quad \phi_{ACC} = (16.62 \pm 23.87)^\circ \quad (10.4)$$

2. A sorpresa invece, le distribuzioni dell'angolo  $\phi_{2-3}$  coincidono. Questo fenomeno è particolare poiché le coppie di basi che interagiscono sono sempre diverse in ogni sequenza (A-T, A-C, C-C e le rispettive speculari, quindi T-A, T-G e G-G);

$$\phi_{AAT}^{2-3} = (10.02 \pm 7.06)^\circ \quad \phi_{AAC}^{2-3} = (10.021 \pm 7.52)^\circ \quad \phi_{ACC}^{2-3} = (10.16 \pm 6.60)^\circ \quad (10.5)$$

3. Il diagramma della distribuzione degli angoli  $\phi_{3-4}$  invece ci conferma quanto aspettato. In questo caso è chiara la maggiore probabilità della molecola ACC di posizionarsi ad angolazioni più elevate delle altre due, mentre la AAT risulta essere la sequenza meno piegata tra le tre;

$$\phi_{AAT}^{3-4} = (15.59 \pm 7.34)^\circ \quad \phi_{AAC}^{3-4} = (17.19 \pm 7.09)^\circ \quad \phi_{ACC}^{3-4} = (18.91 \pm 7.32)^\circ \quad (10.6)$$

4. Le tre tipologie di sequenze, pur essendo accomunate anche nelle basi successive al gruppo AATT, cominciano a comportarsi in modo differente rispetto all'angolazione della coppia G-A e T-C che viene registrata dall'angolo  $\phi_{4-5}$ . In questo caso cominciamo a notare come vi sia un leggero discostamento verso piegature maggiori della sequenza ACC. Questo fenomeno dipende certamente dal fatto che muovendoci verso gli estremi del polimero le basi cominciano a sentire interazioni differenti a causa della differenza delle basi estremali;

$$\phi_{AAT}^{4-5} = (12.13 \pm 6.02)^\circ \quad \phi_{AAC}^{4-5} = (12.25 \pm 6.28)^\circ \quad \phi_{ACC}^{4-5} = (12.60 \pm 5.98)^\circ \quad (10.7)$$

5. Spostandoci verso il cuore della molecola le basi centrali risultano sostanzialmente sempre poco piegate e la loro distribuzione è grosso modo uguale per i tre tipi di DNA. Questo perché il cuore di tutte le sequenze esaminate è il medesimo tratto AATT i cui angoli sono gli  $\phi_{5-6}$  e  $\phi_{6-7}$  e la loro interazione risente in maniera trascurabile della presenza di differenti basi agli estremi.

$$\phi_{AAT}^{5-6} = (8.20 \pm 4.51)^\circ \quad \phi_{AAC}^{5-6} = (8.58 \pm 4.83)^\circ \quad \phi_{ACC}^{5-6} = (8.69 \pm 4.84)^\circ \quad (10.8)$$

$$\phi_{AAT}^{6-7} = (7.51 \pm 4.29)^\circ \quad \phi_{AAC}^{6-7} = (7.53 \pm 4.19)^\circ \quad \phi_{ACC}^{6-7} = (7.18 \pm 4.03)^\circ \quad (10.9)$$

Osserviamo che la scelta fatta durante la definizione delle osservabili di rimuovere dall'analisi i contributi legati alla prima ed all'ultima coppia di basi è ragionevole, in quanto affette da una marcato effetto di disaccoppiamento delle stesse dovuto al fenomeno del "fraying".

Questa prima osservabile ci ha fornito un primo risultato incoraggiante, il fenomeno della differente piegatura sembra essere strettamente correlato alla composizione chimico-strutturale di cui la molecola è composta.

In secondo luogo sarà interessante constatare come questo fenomeno si riporti nelle distanze tra le coppie di basi. E' ragionevole supporre, anche alla luce di quanto visto per le densità del sistema, che le distanze tra le basi aumentino all'aumentare del grado di piegatura. Questo perché le energie di stacking responsabili dell'impilamento difficilmente permettono alle coppie di basi si comprimersi, avvicinandosi l'un l'altra, ulteriormente. In questo modo si favorisce il fenomeno dello "sliding" tra le coppie che scorrono tra loro aumentando la distanza tra i baricentri delle stesse.

Vedremo che i diagrammi delle osservabili legate alla distanza baricentrale tra le basi ci daranno esattamente questa indicazione. Anche in questo caso nei grafici riproponiamo lo schema di appaiamento proposto in precedenza:

$$D_{1-2} = D_{11-12} \quad D_{2-3} = D_{10-11} \quad D_{3-4} = D_{9-10} \quad (10.10)$$

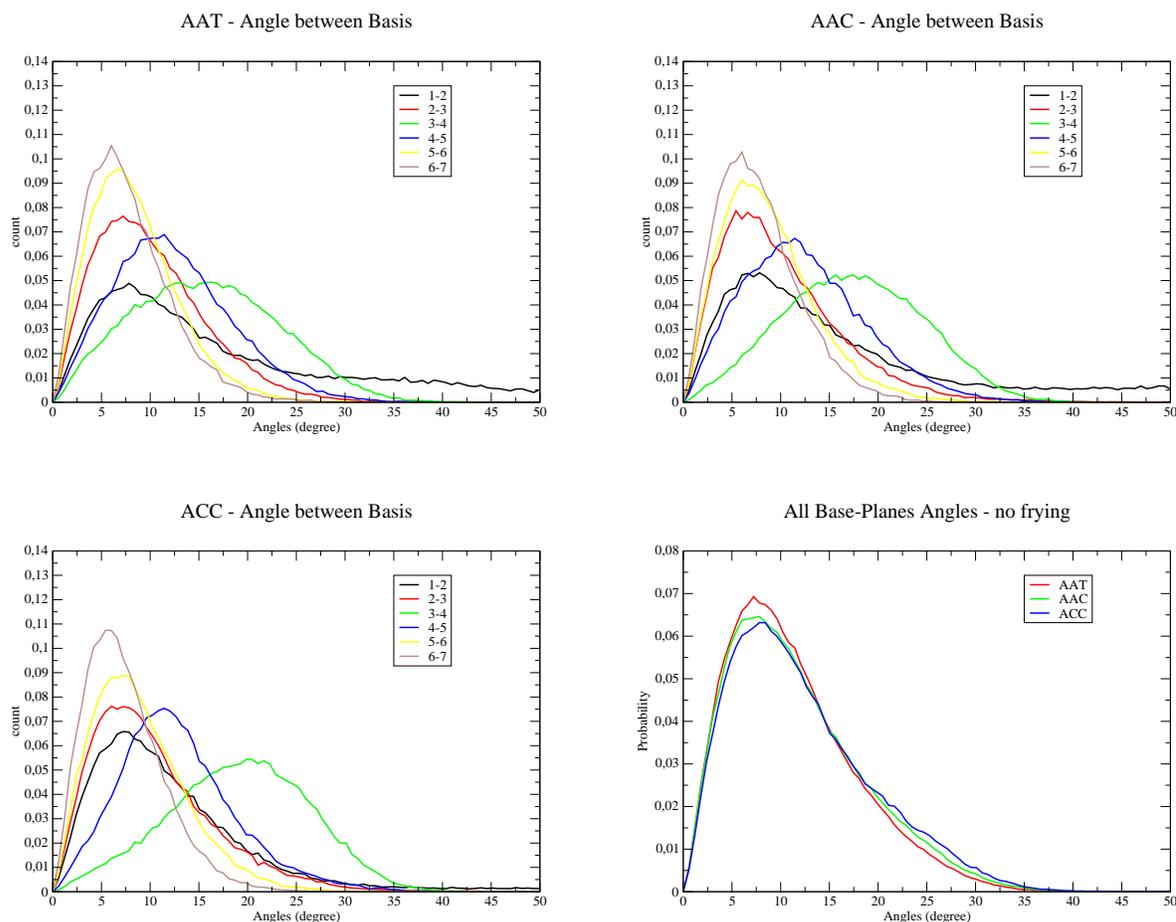


Figura 10.1: Nei grafici vengono riportate tutte le distribuzioni angolari in funzione del tipo di sequenza. Nell'ultimo in basso a destra invece c'è il confronto diretto della somma di tutte le configurazioni per ciascuna delle coppie di nucleotidi che costituiscono la sequenza.

$$D_{4-5} = D_{8-9} \quad D_{5-6} = D_{7-8} \quad D_{6-7} = D_{7-6} \quad (10.11)$$

dove  $D$  è la distanza tra i baricentri delle basi ed ovviamente la  $D_{6-7}$  è determinata da un solo contributo.

In figura 10.3 vengono riportate le distribuzioni di tutte le distanze per ciascuna sequenza mentre nei grafici di figura 10.4 osserviamo nel dettaglio come le singole distanze si modificano al variare della sequenza:

1. Nel primo grafico, che riporta la distribuzione di  $D_{1-2}$ , risulta ancora evidente la presenza del “fraying” tra le basi estremali e giustifica ancora una volta la scelta di eliminare le ultime basi anche nel calcolo della distanza “end-to-end”;
2. Il secondo grafico ci mostra una cosa non aspettata, ovvero la distanza tra le basi risulta

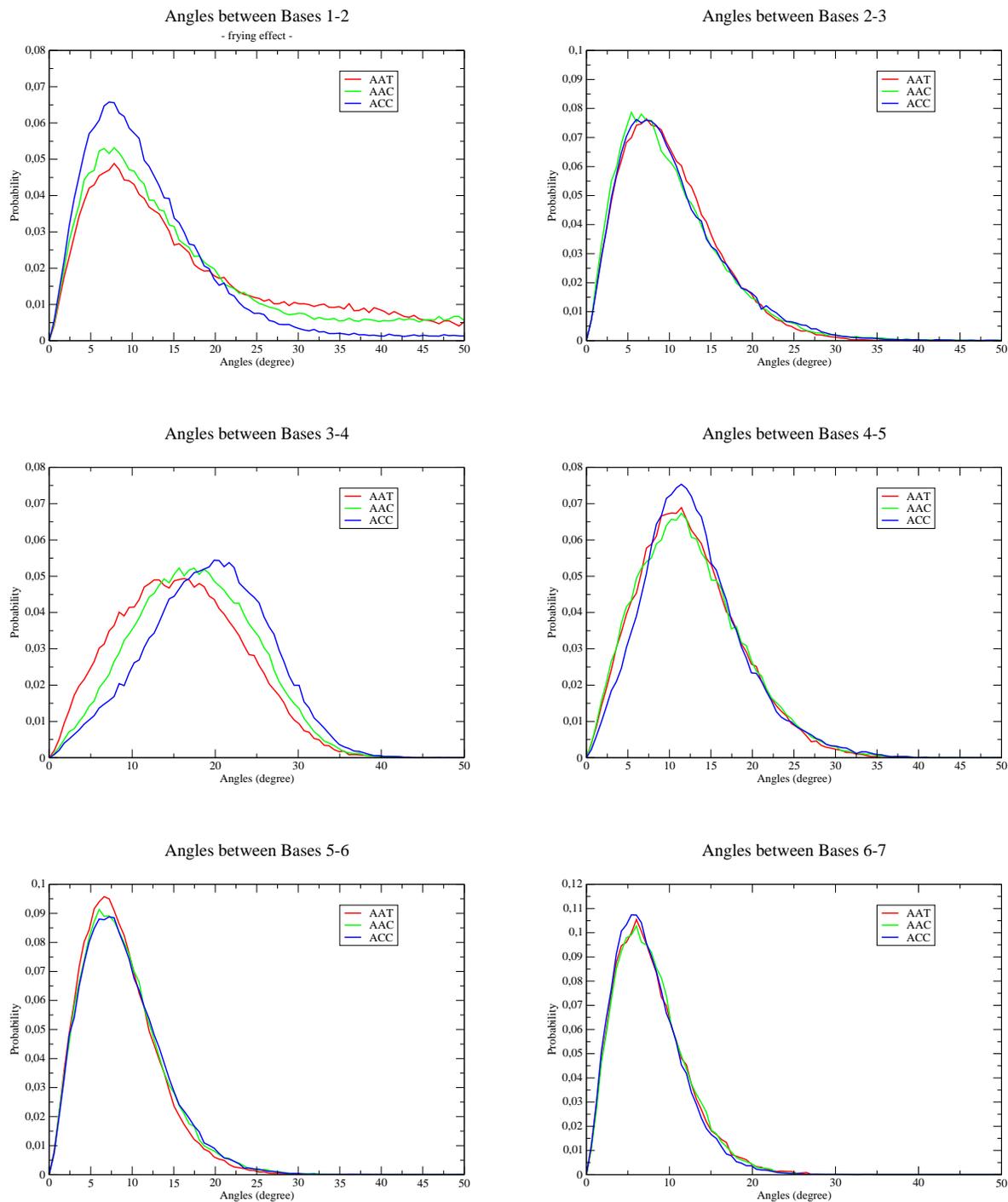


Figura 10.2: grafici delle distribuzioni angolari relative alle medesime coppie di piani che approssimano la coppia di basi in funzione della sequenza considerata.

maggiore nella AAT e nella ACC e minore nella AAC. Questo potrebbe essere dovuto ad un aumento del “rising”, ovvero l’altezza tra le basi sembrerebbe variare in questo modo;

3. La distanza  $D_{3-4}$  sembra essere allineata con il ragionamento fatto in precedenza, in cui la distanza tra le basi deve aumentare nel caso in cui la molecola viene compressa e piegata. Notiamo inoltre un picco secondario nella distribuzione della AAT, probabilmente una delle diciotto sequenze si è trovata in condizioni di equilibrio morfologico differenti dalle altre;
4. Stavolta, nel calcolare la distanza baricentrale  $D_{4-5}$ , è la sequenza ACC ad avere un picco secondario a distanze maggiori, per il resto le tre distribuzioni sono sostanzialmente uguali;
5. Anche negli spostamenti  $D_{5-6}$  e  $D_{6-7}$  delle coppie più interne i tre tipi di sequenze si comportano nello stesso modo, peraltro congruentemente con quanto visto nel caso precedente.

Anche da questo punto di vista la nostra analisi sembra andare nella direzione giusta ma dovremo vedere quanto pesa il diverso contributo della distanza  $D_{2-3}$  nel piegamento effettivo delle sequenze.

A questo punto abbiamo analizzato nel dettaglio i comportamenti a coppie dei nucleotidi di un sistema a diciotto DNA, ottenendo indicazioni confortanti per quanto riguarda l’ordinamento del grado di piegatura aspettato. Occorre però astrarre da questi valori ottenuti ed ottenere delle stime concrete di quanto effettivamente piega la molecola nel complesso, cosa che faremo attraverso l’introduzione delle nuove osservabili.

## 10.2 Bending assiale

Siamo finalmente arrivati ad analizzare l’osservabile che ci consentirà di quantificare per la prima volta il bending delle sequenze. Alla luce della definizione ripresa dal testo di Dickerson [28] e modificata nell’analisi considerando i baricentri invece degli atomi del backbone, riportiamo in un grafico il calcolo di tutti gli angoli di ciascuna delle diciotto sequenze contenute nel box.

Come nel caso precedente tale grafico, ottenuto mediante l’analisi con xmgrace di tutte i valori degli angoli trasformati in un istogramma che è stato successivamente interpolato, rappresenta la distribuzione di probabilità che una particolare tipologia di sequenza assuma un dato angolo all’interno del sistema.

Come previsto nel grafico in figura 10.5 di tutti gli angoli ottenuti mostra con chiarezza che le sequenze piegano in maniera molto differente. La AAT risulta la sequenza che con

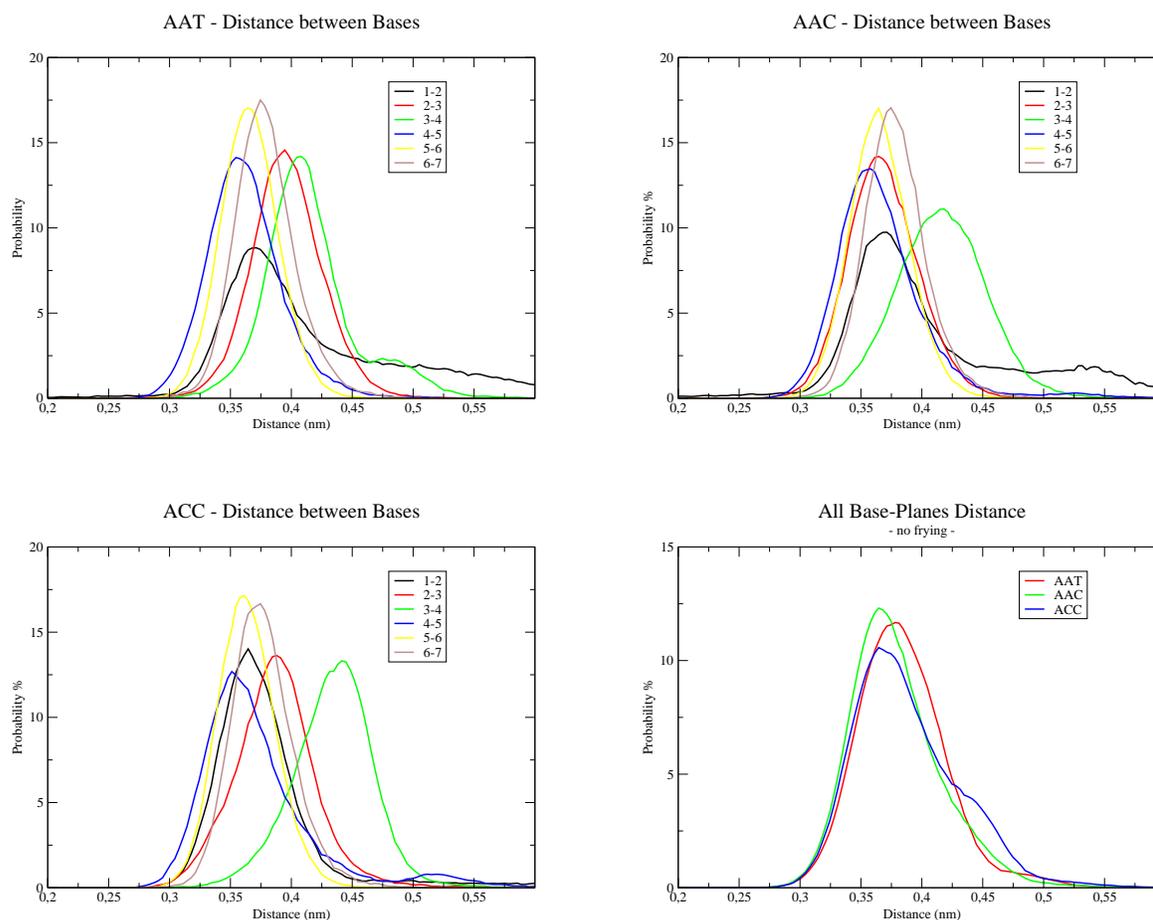


Figura 10.3: Diagrammi di tutte le distanze tra coppie di basi azotate per ciascuno dei tipi di sequenza considerati.

più probabilità assume conformazione rettilinea, subito seguita dalla AAC che sembra piegare leggermente meno. Tra le tre la ACC è chiaramente la più piegata, preferendo molto frequentemente conformazioni ad angolarità minore.

Per ottenere la stima dell'angolo medio e per vedere che il sistema è effettivamente stabilizzato dal punto di vista morfologico è necessario mostrare anche l'evoluzione temporale della media delle configurazioni a diversa piegatura. Decidiamo di operare in questo modo: poiché abbiamo a disposizione diciotto traiettorie differenti possiamo pensare di considerarle tra loro indipendenti e di calcolare una traiettoria media che rappresenta l'evoluzione globale del sistema. Sarà di questa traiettoria media che calcoleremo il bending assiale e la media al variare del tempo, considerando sempre 100ns come il tempo necessario all'equilibratura del sistema.

In figura 10.6 è riportato l'andamento della media temporale della traiettoria che incorpora

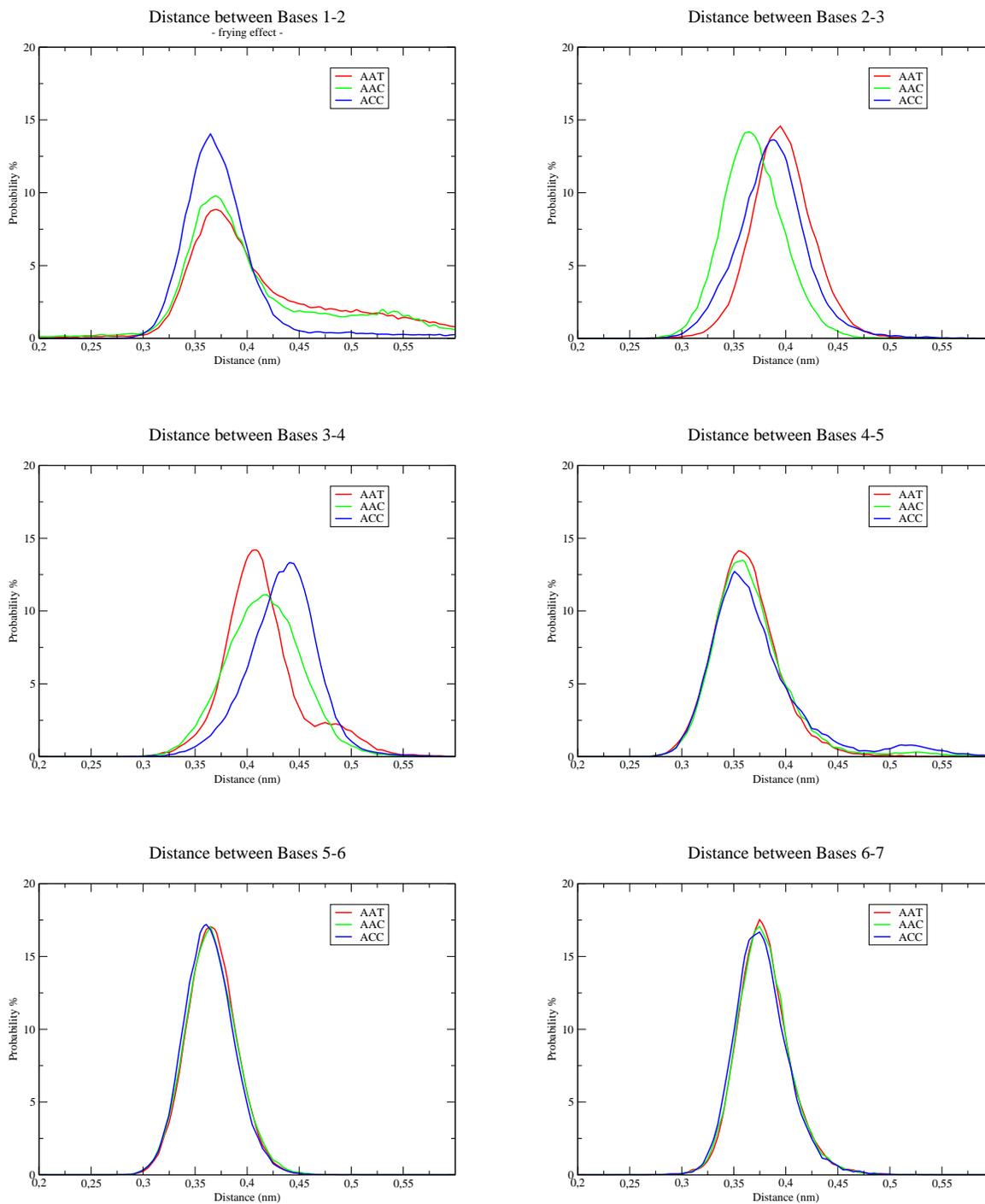


Figura 10.4: Distribuzioni delle distanze baricentriche di coppie di nucleotidi successive per tutte le sequenze studiate.

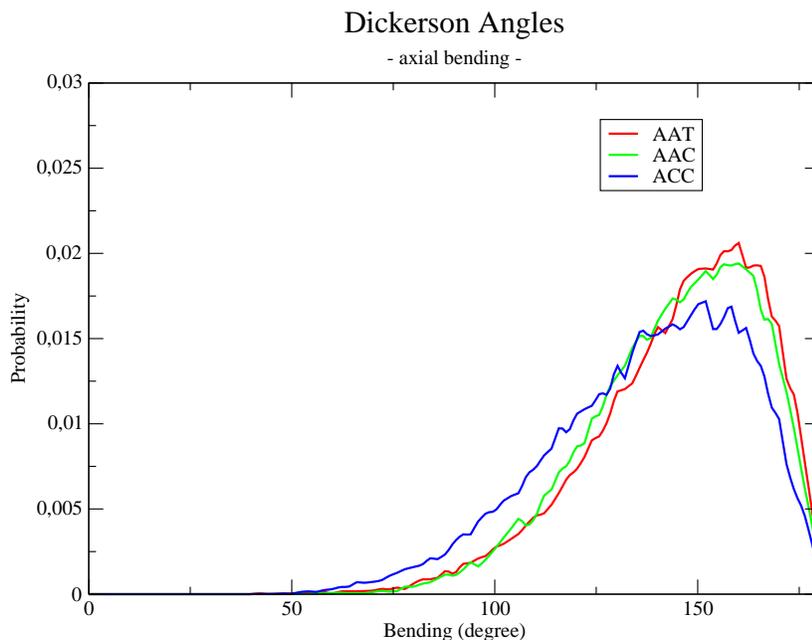


Figura 10.5: Distribuzione degli angoli di Dickerson osservati per tutte le sequenze delle simulazioni a molti DNA.

tutte le diciotto. Notiamo che l'andamento è esattamente quello previsto dal nostro modello e che in particolare per tutta la durata delle simulazioni le sequenze AAT e AAC risultano avere angoli pressapoco simili mentre la ACC è decisamente la più piegata tra le due.

L'analisi statistica dei risultati ci porta a considerare i seguenti valori per il bending del sistema:

$$\Theta_{AAT} = (144.93 \pm 3.18)^\circ \quad (10.12)$$

$$\Theta_{AAC} = (143.07 \pm 3.42)^\circ \quad (10.13)$$

$$\Theta_{ACC} = (136.43 \pm 3.33)^\circ. \quad (10.14)$$

E' doveroso ricordare che questi valori sono delle stime dei valori di piegatura globali, infatti nel calcolarli non stiamo considerando le ultime coppie di basi per colpa del "fraying". Quando il sistema reale nematizza, invece, anche gli ultimi nucleotidi concorrono nella formazione della fase liquido cristallina fornendo quindi un contributo non trascurabile al grado di piegatura di ciascuna delle sequenze.

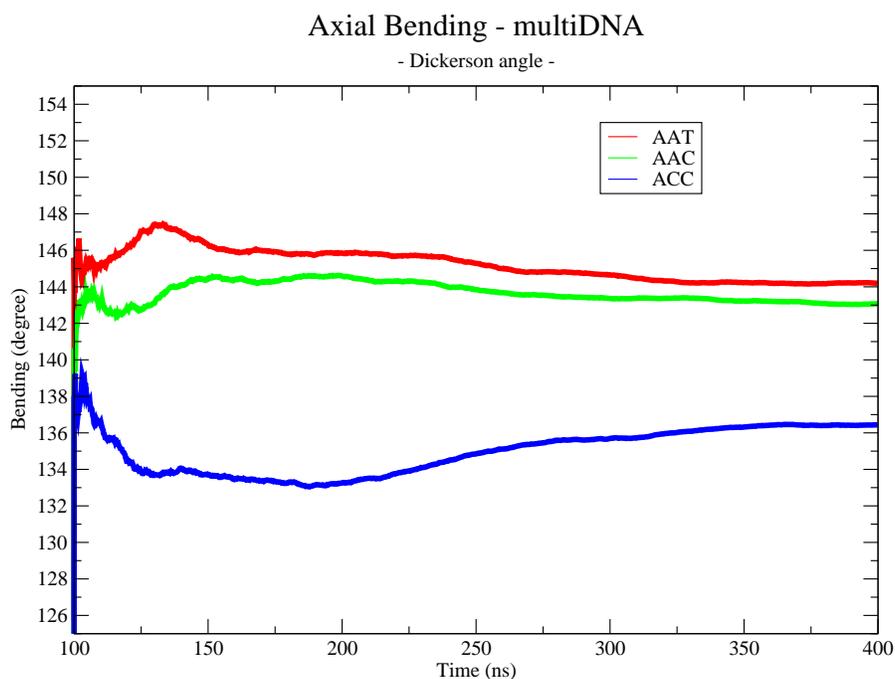


Figura 10.6: Bending assiale calcolato con il metodo di Dickerson e mediato dopo 100ns di equilibratura.

### 10.3 “Bending” fosforico

Abbiamo a disposizione un ulteriore strumento per valutare il bending strutturale dell’acido nucleico, quello legato agli angoli formati tra il baricentro del gruppo centrale e i baricentri dei gruppi fosforici terminali. Come già detto questo contributo ci consente di spingerci più in là nell’esaminare la struttura permettendoci di arrivare a ridosso delle ultime coppie di basi azotate.

Analogamente a quanto fatto nel caso del bending assiale, riportiamo in un istogramma interpolato la distribuzione angolare di ciascuna configurazione temporale di piegatura. Così facendo otteniamo nuovamente quanto previsto, ovvero le distribuzioni risultano più piccate in corrispondenza di angoli maggiori per quella che deve essere la sequenza più dritta, la AAT, poco distante troviamo la AAC mentre il picco della ACC è visibilmente più spostato verso piegature maggiori.

Per poter apprezzare il fenomeno anche nel tempo possiamo anche qui calcolare il bending fosforico servendoci della configurazione media delle diciotto traiettorie di ciascun sistema. In figura 10.8 possiamo vedere l’andamento temporale della media dell’osservabile e constatare anche stavolta che otteniamo quanto previsto: l’ordine di piegatura è mantenuto correttamente a tutti i tempi.

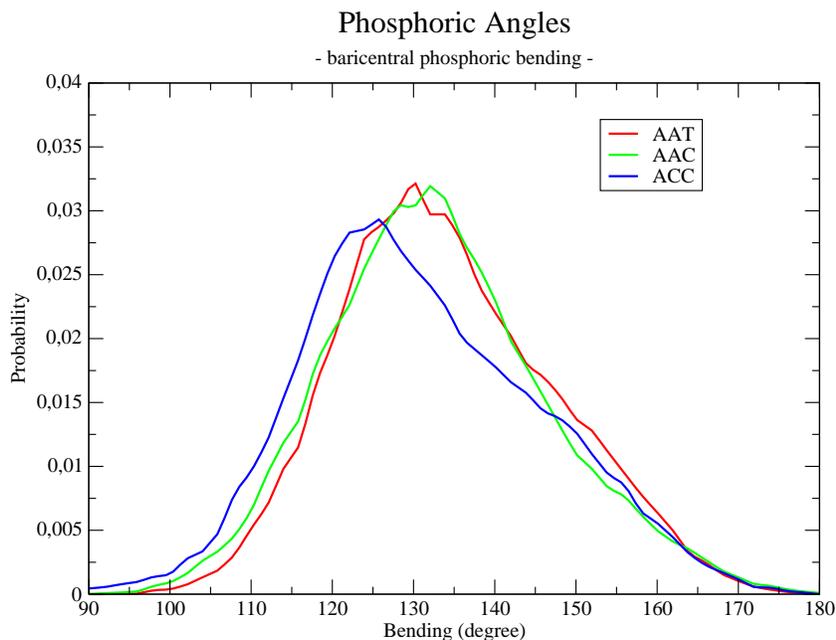


Figura 10.7: Distribuzione degli angoli di piegatura fosforica riscontrati per tutte le diciotto sequenze per ciascuno dei tre sistemi. In questo caso è evidente lo spostamento verso configurazioni più piegate della ACC.

C'è da dire però che questo osservabile non risulta stabile come quello definito da Dickerson, infatti specialmente verso la parte finale della simulazione le medie della sequenza ACC sembrano risalire. Nonostante questo fatto il sistema mantiene delle conformazioni a piegatura ben definite lungo tutta la durata della simulazione, di ben  $400ns$ , rispecchiando la situazione prevista dalle ipotesi fatte sulla concentrazione di nematizzazione.

I valori che utilizzeremo nel seguito per la creazione del nostro modello a doppio cilindro piegato verranno estrapolati direttamente da quanto osservato in via computazionale. Le medie dei valori del “bending” di ciascuna struttura forniscono i valori riportati di seguito:

$$\theta_{AAT} = (133.23 \pm 2.73)^\circ \quad (10.15)$$

$$\theta_{AAC} = (132.01 \pm 2.75)^\circ \quad (10.16)$$

$$\theta_{ACC} = (130.17 \pm 2.82)^\circ. \quad (10.17)$$

Poiché per come abbiamo definito il modello a doppio cilindro occorre anche calcolare la distanza “end-to-end” media delle strutture molecolari, risulta interessante lo studio della media temporale di tale distanza. Il grafico 10.9 mostra chiaramente come questo osservabile

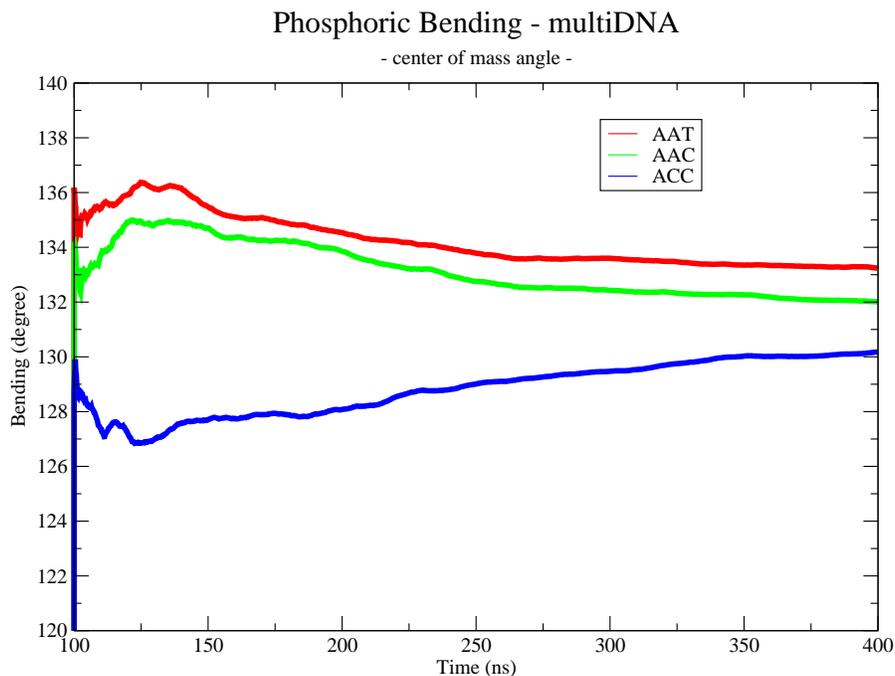


Figura 10.8: Media del grado di “bending” calcolato con il metodo delle congiungenti dei baricentri dei gruppi fosforici. All’aumentare della durata della simulazione i tre valori sembrano convergere, pur mantenendo sempre il corretto ordinamento.

sia molto più regolare degli altri precedentemente esaminati.

Come previsto anche in questo caso riscontriamo un andamento consistente con quanto ottenuto da Nakata, ed anche la distanza “end-to-end” può giocare un ruolo cruciale nella determinazione della concentrazione critica a cui il sistema transisce. Effettuando un’analisi statistica ottengo per tale distanza i valori:

$$d_{AAT} = (35.928 \pm 2.337) \text{ \AA} \quad (10.18)$$

$$d_{AAC} = (35.384 \pm 2.557) \text{ \AA} \quad (10.19)$$

$$d_{AAT} = (33.912 \pm 2.282) \text{ \AA} \quad (10.20)$$

Grazie alle osservabili definite con i gruppi fosforici abbiamo ottenuto ben due informazioni:

- Per prima cosa il fenomeno di piegatura è criticamente dipendente dalla sequenza considerata, pur se le differenze in termini di angoli sono molto piccole, dell’ordine del 3% tale comportamento è sempre ben visibile;
- Anche l’elongazione dipende in modo cruciale dal tipo di acido nucleico che stiamo

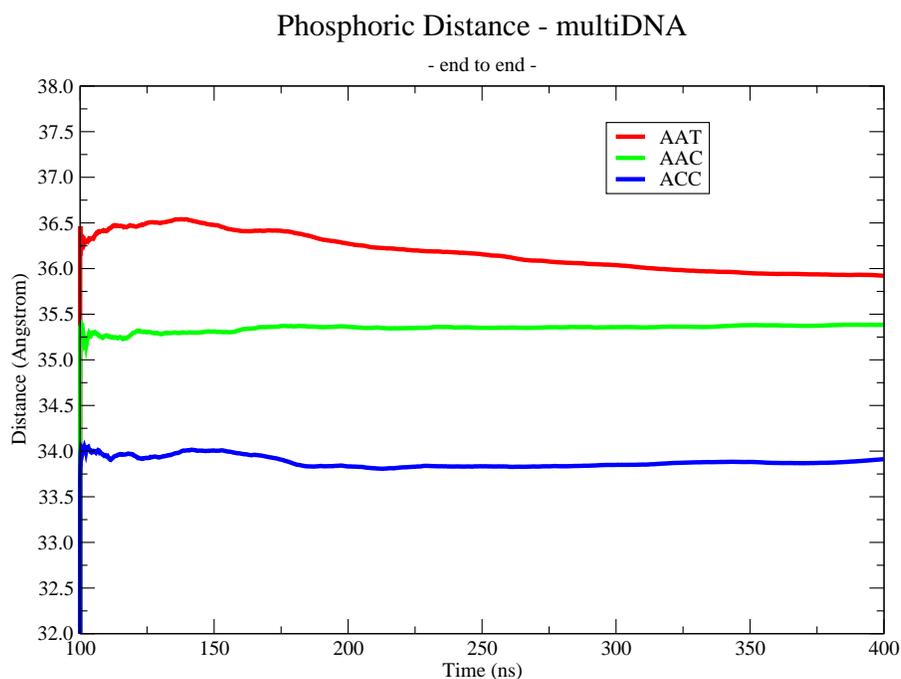


Figura 10.9: Plottaggio dell'andamento dei valori medi in funzione del tempo iniziando il conteggio dopo  $100ns$  di equilibratura iniziale.

trattando e quindi anch'essa sembra correlata alla concentrazione critica a cui avviene la transizione di fase nematica.

Perciò anche se le simulazioni iniziano tutte da una configurazione cristallografica, che ha sostanzialmente le stesse proprietà geometriche indipendentemente dalla sua composizione chimica, essa nel tempo evolve per arrivare a stabilizzarsi in strutture sostanzialmente modificate dal punto di vista morfologico. La piegatura e l'elongazione "end-to-end" riscontrate nelle nostre simulazioni sicuramente concorrono in modo rilevante a determinare la differenza di densità dei sistemi, già notata preventivamente in fase di equilibratura, ed è quindi ragionevole supporre come tale modifica di forma della molecola rivesta un ruolo di particolare importanza nella transizione di fase liquido cristallina osservata sperimentalmente.

## Parte III

# Simulazione Monte Carlo di cilindri piegati

## Capitolo 11

# Teoria sulla dinamica Monte Carlo

Abbiamo già discusso del fatto che la dinamica molecolare non sia in grado di riprodurre correttamente il comportamento termodinamico di un sistema molto complesso come quello studiato precedentemente. Infatti anche nelle simulazioni di Maffeo [21], effettuate facendo evolvere 458 segmenti corti di DNA in soluzione acquosa ionizzata, non è stata riscontrata la nematizzazione del sistema benché si osservasse la formazione di aggregati lineari dovuti alle forze di “stacking” idrofobico. Per giunta la potenza computazionale richiesta è molto elevata e prevedere il comportamento di un sistema dinamico molto grande richiede non solamente grandi risorse di calcolo ma anche tempi di simulazione piuttosto lunghi.

Per sondare le caratteristiche termodinamiche del sistema è necessario quindi procedere con un nuovo approccio di natura statistica adatto a modellizzare il sistema preso in esame, attraverso l'introduzione di oggetti che approssimano le strutture degli acidi nucleici appena studiati. Sappiamo che un approccio del genere ha fornito, mediante il modello OBF [14, 15], una spiegazione ragionevole del fatto che lunghi segmenti di DNA rettilinei transivano in una fase nematica a determinate densità tipiche della transizione, grazie a contributi di volume escluso. Tale modello utilizza la dinamica Monte Carlo (MC) per esplorare lo spazio delle fasi di un sistema formato da oggetti a geometria cilindrica caratterizzati esclusivamente dal rapporto di forma  $X_0 = \frac{L}{D}$  (“aspect-ratio”) che ne definisce le proporzioni, prevedendo un limite inferiore al di sotto della quale il sistema, per soli contributi di volume escluso, non riusciva più a transire nella fase nematica.

Quando è stato portato alla luce che il fenomeno di nematizzazione si presentava anche per segmenti di DNA composti da 6 a 20 basi azotate [17, 16] è stata necessaria una revisione del vecchio modello. La ricomparsa del fenomeno è stata addebitata alla presenza di interazioni di “stacking” idrofobico che intercorrono tra le basi terminali dei segmenti di acido nucleico in soluzione, e la modellizzazione del fenomeno, dal punto di vista MC, è stata possibile mediante l'aggiunta di due siti attrattivi posti sulle basi del cilindro che approssima la struttura a doppia elica. In questo modo le interazioni di “hard-core” sono state affiancate da vere interazioni

attrattive ed è stato possibile nuovamente osservare la nematizzazione del sistema [19].

Tale modello proposto, però, è ancora caratterizzato esclusivamente dalla grandezza  $X_0$  e perciò dipende esclusivamente dal numero di basi da cui il biopolimero è costituito. Questo ancora non rappresenta in modo efficace la situazione reale poiché sperimentalmente in [16] viene osservato che la transizione di fase dei sistemi composti da dodecameri<sup>1</sup> di DNA avviene a differenti concentrazioni, dipendenti criticamente dalla sequenza presa in esame, con variazioni della  $c_N$  che arrivano fino al 70%.

Alla luce di ciò viene naturale proporre una modifica al modello mediante l'introduzione di una nuova grandezza che caratterizza la transizione nematica. E' ragionevole supporre che una differente piegatura dei segmenti corti di DNA influisca in modo cruciale nel determinare le concentrazioni a cui il sistema transisce: un sistema composto da oggetti più piegati infatti, avrà bisogno di raggiungere concentrazioni più elevate per poter nematizzare.

Da quanto visto sperimentalmente [16] per segmenti di B-DNA lunghi 12 basi azotate, e considerando valida la nostra ipotesi, la sequenza ACC deve risultare in media più piegata delle altre, seguita dalla AAC e dalla AAT che dovranno essere più rettilinee. Questo fenomeno è stato evidenziato per via computazionale mediante simulazioni a dinamica molecolare nel capitolo precedente, ottenendo in particolare per le simulazioni multi-DNA, l'andamento previsto per tutta la durata della simulazione.

Per determinare come questo giochi un ruolo nella termodinamica è necessario passare ad un approccio statistico attraverso simulazioni MC. Il modello proposto nel seguito, quindi, propone di far uso dello stesso cilindro con siti attrattivi posti sulle basi terminali precedentemente utilizzato modificandolo attraverso la piegatura rispetto al suo asse principale di un angolo determinato in base al tipo di sequenza, per poi successivamente caratterizzarne il comportamento termodinamico collettivo prestando particolare attenzione alla formazione delle fasi.

L'ipotesi che tenteremo di verificare è che all'aumentare della piegatura del sistema la transizione nematica si sposterà verso densità più elevate, correlando con le misure ottenute sperimentalmente, e che tale variazione della  $\varphi_N$  riesca a confrontarsi con la grande variazione ottenuta sperimentalmente.

Per fare questo ci serviremo della simulazione MC di un sistema composto da  $N = 500$  oggetti formati dall'unione di due cilindri piegati tra loro rispetto a ciascun asse considerati come un unico corpo rigido, supponendo fissate la temperatura  $T$  e pressione  $P$ , per ottenere il diagramma di fase che caratterizzerà la transizione lavorando in un ensemble statistico di tipo NPT e da questo stimare la "volume fraction" di nematizzazione.

---

<sup>1</sup>Per quanto riguarda acidi nucleici formati da 14 basi azotate invece le concentrazioni critiche sono sostanzialmente indipendenti dalla sequenza, pari a circa  $400^{mg/ml}$ .

## 11.1 Simulazioni Monte Carlo NVT

Come primo approccio alla teoria della simulazione Monte Carlo decidiamo di analizzare [5] l'insieme statistico più semplice con un numero fissato di particelle ( $N$ ) in un determinato volume ( $V$ ) ad una temperatura costante controllata esternamente ( $T$ ): per brevità ci riferiremo quindi ad un ensemble di tipo  $NVT$ , chiamato anche "ensemble canonico".

Supponendo il nostro generico sistema descritto da una Hamiltoniana separabile in due contributi  $\mathcal{H}(\mathbf{p}^N, \mathbf{r}^N) = \mathcal{K}(\mathbf{p}^N) + \mathcal{U}(\mathbf{r}^N)$  possiamo scrivere l'espressione per la funzione di partizione che caratterizza l'ensemble:

$$Q = c \int d\mathbf{p}^N d\mathbf{r}^N e^{-\frac{\mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}{k_B T}}. \quad (11.1)$$

Grazie a questa quantità è possibile calcolare le grandezze macroscopiche del sistema  $A$  come medie sull'ensemble statistico:

$$\langle A \rangle = \frac{\int d\mathbf{p}^N d\mathbf{r}^N A(\mathbf{p}^N, \mathbf{r}^N) e^{-\beta \mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}}{\int d\mathbf{p}^N d\mathbf{r}^N e^{-\beta \mathcal{H}(\mathbf{p}^N, \mathbf{r}^N)}} \quad (11.2)$$

In generale è semplice calcolare la parte di integrazione dipendente dai momenti  $\mathbf{p}$  mentre la difficoltà computazionale sta nella dipendenza dalla posizione  $\mathbf{r}$ . Per questo motivo introduciamo il metodo Monte Carlo nel calcolo della grandezza fisica osservabile  $\langle A \rangle$ :

$$\langle A \rangle = \frac{\int d\mathbf{r}^N A(\mathbf{r}^N) e^{-\beta \mathcal{U}(\mathbf{r}^N)}}{\int d\mathbf{r}^N e^{-\beta \mathcal{U}(\mathbf{r}^N)}} \quad (11.3)$$

Per prima cosa scriviamo la parte della funzione di partizione dipendente dai momenti:

$$Z = \int d\mathbf{r}^N e^{-\beta \mathcal{U}(\mathbf{r}^N)} \quad (11.4)$$

definendo la densità di probabilità normalizzata come:

$$\mathcal{N}(\mathbf{r}^N) = \frac{e^{-\beta \mathcal{U}(\mathbf{r}^N)}}{Z}. \quad (11.5)$$

Ora supponiamo di essere in grado di generare dei punti nella configurazione spaziale in accordo con questa distribuzione di probabilità. Questo comporta che in media, il numero di punti generati per unità di volume  $n_i$  attorno al punto  $\mathbf{r}^N$  è uguale a  $L\mathcal{N}(\mathbf{r}^N)$ , dove con  $L$  abbiamo inteso il numero totale di punti generati secondo la distribuzione di probabilità. A

questo punto l'integrale può essere discretizzato ottenendo così:

$$\langle A \rangle = \frac{1}{L} \sum_{i=1}^L A(\mathbf{r}_i^N). \quad (11.6)$$

Dalla meccanica statistica sappiamo che la distribuzione di probabilità di un ensemble è fortemente piccata in una piccolissima regione dello spazio delle fasi rispetto al suo volume totale. In queste condizioni l'applicazione diretta del metodo MC non aiuta. Dobbiamo perciò trovare uno schema logico con cui esplorare esclusivamente le posizioni rilevanti del sistema, ovvero quelle configurazioni microscopiche del sistema in cui la distribuzione di probabilità è non nulla.

L'intento è che il nostro integrale venga valutato effettuando un "random walk" pesato sulla distribuzione nello spazio delle fasi  $\mathcal{N}(\mathbf{r}^N)$ , vogliamo cioè effettuare quello che viene chiamato *campionamento per importanza*. Considerando la probabilità di transizione da uno stato  $o$  ad uno stato  $n$  generato casualmente, chiamandola  $\pi(o \rightarrow n)$  è possibile mostrare [5], in regime stazionario indipendente dal tempo, che il principio del bilancio dettagliato:

$$\mathcal{N}(o) \pi(o \rightarrow n) = \mathcal{N}(n) \pi(n \rightarrow o). \quad (11.7)$$

è una condizione sufficiente ad effettuare un "random walk" pesato nello spazio delle fasi.

Esistono moltissimi modi per scegliere una matrice di transizione che soddisfi la 11.7. Cominciamo col considerarla come il prodotto di una probabilità di effettuare un movimento da  $o$  ad  $n$ , che indichiamo con  $\alpha(o \rightarrow n)$ , ed una probabilità di accettare questo movimento  $acc(o \rightarrow n)$  compatibilmente con la distribuzione di probabilità nell'ensemble considerato. Grazie all'introduzione di queste due grandezze possiamo scrivere:

$$\pi(o \rightarrow n) = \alpha(o \rightarrow n) acc(o \rightarrow n). \quad (11.8)$$

Nello schema originale proposto da Metropolis, la matrice di transizione è simmetrica ovvero  $\alpha(o \rightarrow n) = \alpha(n \rightarrow o)$ , il che implica per il bilancio dettagliato:

$$\mathcal{N}(o) acc(o \rightarrow n) = \mathcal{N}(n) acc(n \rightarrow o) \quad (11.9)$$

che ci consente di ricavare il rapporto tra le probabilità di accettazione della mossa casuale:

$$\frac{acc(o \rightarrow n)}{acc(n \rightarrow o)} = \frac{\mathcal{N}(n)}{\mathcal{N}(o)} = e^{-\beta[\mathcal{U}(n) - \mathcal{U}(o)]}. \quad (11.10)$$

Anche in questo caso esistono molte soluzioni di questa equazione, perciò analizziamo

quella scelta da Metropolis che si è dimostrata molto efficiente in termini computazionali:

$$\begin{aligned} \text{acc}(o \rightarrow n) &= \frac{\mathcal{N}(n)}{\mathcal{N}(o)} \quad \text{se } \mathcal{N}(n) < \mathcal{N}(o) \\ &= 1 \quad \text{se } \mathcal{N}(n) < \mathcal{N}(o) \end{aligned} \quad (11.11)$$

questo comporta per la probabilità di transizione dallo stato  $o$  allo stato  $n$  la soluzione:

$$\begin{aligned} \pi(o \rightarrow n) &= \alpha(o \rightarrow n) \frac{\mathcal{N}(n)}{\mathcal{N}(o)} \quad \text{se } \mathcal{N}(n) < \mathcal{N}(o) \\ &= \alpha(o \rightarrow n) \quad \text{se } \mathcal{N}(n) < \mathcal{N}(o) \\ \pi(o \rightarrow o) &= 1 - \sum_{n \neq o} \pi(o \rightarrow n) \end{aligned} \quad (11.12)$$

Ora esaminiamo il modo con cui un movimento casuale può essere o meno accettato. Supponiamo di aver generato un movimento e di esserci portati dallo stato  $o$  allo stato  $n$  con  $\mathcal{U}(n) > \mathcal{U}(o)$ . In accordo con l'equazione 11.10 la mossa casuale viene accettata con una probabilità pari a:

$$\text{acc}(o \rightarrow n) = e^{-\beta[\mathcal{U}(n) - \mathcal{U}(o)]} < 1 \quad (11.13)$$

Per decidere se accettare o meno la mossa serviamoci di un numero casuale  $Ranf$  generato partendo da una distribuzione di probabilità uniforme nell'intervallo  $[0, 1]$ . Chiaramente la probabilità che  $Ranf$  sia minore di  $\text{acc}(o \rightarrow n)$  è esattamente uguale ad  $\text{acc}(o \rightarrow n)$ . Questo ci consente di affermare che possiamo accettare la mossa se

$$Ranf < \text{acc}(o \rightarrow n) \quad (11.14)$$

e scartarla nel caso contrario. Questa regola ci garantisce che la probabilità di accettare una mossa casuale da  $o$  ad  $n$  è uguale a  $\text{acc}(o \rightarrow n)$ .

Anche per fare questo abbiamo diverse possibilità di scelta, ma in questo caso continueremo a seguire la logica proposta da Metropolis. A questo punto possiamo implementare tale modello con il seguente schema:

1. Selezioniamo una particella casualmente tra le  $N$  disponibili nel nostro sistema, e calcoliamo la sua energia  $\mathcal{U}(\mathbf{r}^N)$ ;
2. Scegliamo uno spostamento casuale della particella, ovvero  $\mathbf{r}' = \mathbf{r} + \mathbf{\Delta}$ , calcolandone la sua energia nella nuova posizione  $\mathcal{U}(\mathbf{r}'^N)$ ;
3. Accettiamo la mossa  $\mathbf{r} \rightarrow \mathbf{r}'$  con una probabilità:

$$\text{acc}(o \rightarrow n) = \min \left( 1, e^{-\beta[\mathcal{U}(n) - \mathcal{U}(o)]} \right) \quad (11.15)$$

altrimenti la scartiamo e torniamo alla posizione iniziale, iterando il procedimento dal punto 1. e scegliendo una nuova particella casualmente.

In questo modo abbiamo capito come implementare l'algoritmo di integrazione MC-Metropolis per simulare un sistema di particelle in un ensemble NVT. C'è da dire che spesso questo insieme statistico rappresenta piuttosto male la termodinamica di un esperimento effettuato a laboratorio, dove in generale vengono mantenute costanti la temperatura, il numero di particelle e la pressione  $P$ , lasciando libero di fluttuare il volume all'interno del quale avviene la simulazione. Per questo motivo esaminiamo brevemente come viene strutturata una simulazione a dinamica Monte Carlo nell'ensemble NPT, l'insieme statistico che useremo per effettuare le simulazioni di doppi cilindri piegati.

## 11.2 Simulazioni Monte Carlo NPT

In questa sede siamo interessati principalmente a simulare il comportamento del sistema a pressione e temperatura costante per poter osservare la "volume fraction"  $\varphi_N$  a cui il volume del sistema nematizza, perciò risulta più adeguato effettuare una simulazione Monte Carlo nell'insieme statistico NPT [5], chiamato anche ensemble "isobarico-isotermico". L'utilità di questo metodo non sta solo nel fatto che gran parte degli esperimenti viene eseguito controllando la temperatura e la pressione, ma anche perché tale insieme è particolarmente utile a descrivere una transizione di fase del primo ordine, come quella a cui siamo interessati, nella quale rimangono costanti pressione e la temperatura. Contestualmente a quanto fatto da Glaser [25] e De Michele [19], ci serviremo dell'insieme NPT per ricavare il diagramma di fase di doppi cilindri piegati e, successivamente, individuare la regione in cui avviene la transizione isotropico-nematica.

Per poter fare questo tipo di analisi occorre ovviamente utilizzare una funzione di partizione diversa da quanto visto in precedenza per l'insieme NVT, e di conseguenza adeguare la formula per il calcolo della media delle osservabili.

Analizziamo più nel dettaglio la situazione. Ovviamente in questo ensemble bisogna tener conto delle fluttuazioni di volume del sistema ed è possibile vedere [5] che la funzione di partizione può essere scritta come:

$$Q(N, P, T) = \frac{1}{\Lambda^{3N} N!} \int d(\ln V) ds^N V^{N+1} e^{-\beta PV - \beta U(\mathbf{s}^N, L)} \quad (11.16)$$

nella quale stiamo tenendo conto della riscalatura delle distanze  $\mathbf{r} = L\mathbf{s}$ , dove  $L$  è la lunghezza del lato della scatola.

Inoltre è possibile mostrare che in questo insieme la media di un'osservabile è esprimibile mediante:

$$\langle A \rangle = \frac{1}{Q} \int dV ds^N e^{-\beta PV + N \ln V - \beta U} \quad (11.17)$$

e la probabilità di trovare un volume  $V$  è determinata dalla distribuzione:

$$\mathcal{N}(V, \mathbf{s}^N) \propto V^{N+1} e^{-\beta PV - \beta \mathcal{U}(\mathbf{s}^N, L)}. \quad (11.18)$$

Avendo cambiato ensemble in esame, abbiamo bisogno di un nuovo criterio di accettazione delle mosse casuali con le quali esplorare lo spazio delle fasi. Ora poiché anche il volume può fluttuare, oltre alla variazione delle tre coordinate di una particella scelta casualmente, eseguiremo anche una variazione casuale del volume del sistema. Il criterio di accettazione della mossa appena eseguita è costruito sulla base della distribuzione di probabilità di trovare il sistema in una data configurazione  $\mathcal{N}(V, \mathbf{s}^N)$  e tenendo conto del principio del bilancio dettagliato, ovvero:

$$acc(o \rightarrow n) = \min \left( 1, e^{-\beta [\mathcal{U}(\mathbf{s}^N, V') - \mathcal{U}(\mathbf{s}^N, V) + P(V' - V) - \frac{N+1}{\beta} \ln(\frac{V'}{V})]} \right). \quad (11.19)$$

Una modifica al volume, una volta accettata, richiede il ricalcolo di tutte le interazioni intermolecolari. Questa operazione può essere molto costosa in termini di risorse del sistema e tempi di esecuzione e può essere ottimizzata riscaldando le lunghezze nel caso avessimo scelto il potenziale di Lennard-Jones per governare l'interazione tra HBC. Ma dato che il modello usato da De Michele et al. [19] sfrutta un potenziale di tipo “square well” non è necessario usare l'ottimizzazione suggerita dal Frenkel [5] nemmeno nel nostro caso.

L'algoritmo che utilizzeremo per implementare il metodo di MC-Metropolis nell'insieme NVT è formalmente identico a quello trattato in precedenza, al quale però applichiamo un'accortezza. Per esplorare lo spazio delle fasi non abbiamo bisogno di variare il volume ogni qual volta variamo la posizione di una determinata particella, piuttosto possiamo decidere di variarlo grossomodo ogni volta che abbiamo già variato la posizione di tutte le particelle contenute nella scatola. Questo ci consentirà un'esecuzione più rapida del programma ed un “random walk” più efficiente nello spazio delle fasi.

A questo punto l'algoritmo può essere schematizzato come segue:

1. Prepariamo la configurazione iniziale del sistema fissando le lunghezze dei lati della scatola ed il numero di particelle in essa contenute;
2. Eseguiamo un loop sulle particelle, l'uscita dal ciclo è determinata dall'estrazione di un numero casuale contenuto in un determinato insieme e passando al punto 7;
3. Scegliamo casualmente una particella tra le  $N$  disponibili e decido la sua nuova posizione;
4. Calcoliamo la variazione di energia determinata dallo spostamento della particella considerata e, mediante il metodo di Metropolis utilizzato nell'ensemble NVT, decido se accettare o meno la mossa;

5. Se la mossa è accettata aggiorno la posizione e l'energia del sistema;
6. Scegliamo la prossima particella e torno al punto 2;
7. Si effettua una variazione di volume dell'intero sistema, calcolo il criterio di accettazione e decido se accettare o meno la mossa;
8. Se la mossa è accettata aggiorno la posizione, l'energia ed il volume del sistema e torno al punto 2.

Con questi semplici criteri è possibile implementare la simulazione Monte Carlo per esplorare lo spazio delle fasi di qualsiasi tipo di sistema. Nel nostro caso utilizzeremo il software sviluppato da De Michele per la simulazione di sistemi composti da corpi rigidi di varia natura [27, 19], modificando la forma degli oggetti mediante l'introduzione di una geometria a doppio cilindro piegato che verrà descritta dettagliatamente nel capitolo successivo.

### 11.3 Teoria sulla termodinamica di Cilindri Rigidi con siti attrattivi

Per valutare gli esperimenti computazionali ottenuti dalla simulazione MC di doppi cilindri piegati ("hard bent cylinders" - HBC) decidiamo di confrontare le grandezze ottenute direttamente con quelle dei modelli da cui essi derivano. Prenderemo come teoria di riferimento quella sviluppata in [19, 20] per la termodinamica di cilindri rigidi (HC) e superquadriche di forma quasi-cilindrica (SQ) per una comparazione dei risultati numerici ottenuti.

Assumiamo che l'energia libera del sistema sia descritta dalla seguente relazione:

$$\begin{aligned} \frac{\beta F}{V} = & \sum_{l=1}^{\infty} \nu(l) \{ \ln [v_d \nu(l)] - 1 \} + \frac{\eta(\varphi)}{2} \sum_{l,l'=1}^{\infty} \nu(l) \nu(l') v_{excl}(l, l') + \\ & - (\beta \Delta E_S + \sigma_b) \sum_{l=1}^{\infty} (l-1) \nu(l) + \sum_{l=1}^{\infty} \nu(l) \sigma_o(l) \end{aligned} \quad (11.20)$$

dove  $\nu(l)$  è il numero che descrive la densità della catena di lunghezza  $l$ , normalizzata in modo tale che  $\sum_{l=1}^{\infty} l \nu(l) = \rho$ ,  $v_d$  è il volume del monomero,  $\beta \Delta E_S$  è l'energia di "stacking" tra le basi,  $v_{excl}(l, l')$  è il volume escluso di due catene di lunghezza  $l$  ed  $l'$ ,  $\sigma_b$  è il contributo dell'energia libera perso per formare un legame,  $\eta(\varphi)$  è il fattore di Pearson-Lee pari a:

$$\eta(\varphi) = \frac{1}{4} \frac{4 - 3\varphi}{(1 - \varphi)^2} \quad (11.21)$$

ed infine  $\sigma_o$  tiene conto dell'entropia orientazionale persa da una catena lunga  $l$  nella formazione della fase nematica. Notiamo la separazione del contributo all'energia libera in un

contributo energetico  $\beta\Delta E_S$  sommato ad uno entropico  $\sigma_o$ , e l'inclusione del fattore di Pearson-Lee che sappiamo non funzionare nel caso di catene rigide formate da sfere e ad alte densità. E' possibile trovare una giustificazione più dettagliata delle approssimazioni fatte nell'articolo di De Michele [19].

Secondo la teoria di Van Der Schoot e Cates [57] è possibile assumere per  $v_{excl}(l, l')$  un'espressione polinomiale al secondo ordine in  $l$  ed  $l'$ :

$$v_{excl}(l, l', f(\mathbf{u})) = 2 \int f(\mathbf{u}) f(\mathbf{u}') D^3 \left[ \Psi_1(\gamma, X_0) + \frac{l+l'}{2} \Psi_2(\gamma, X_0) X_0 + \Psi_3(\gamma, X_0) X_0^2 l l' \right] d\Omega d\Omega' \quad (11.22)$$

avendo considerato  $f(\mathbf{u})$  come la probabilità normalizzata per un dato monomero di avere un'orientazione  $\mathbf{u}$  compresa tra  $\Omega$  e  $\Omega + d\Omega$  e  $\Psi_\alpha$  descrive la dipendenza angolare del volume escluso.

Esplicitamente per due catene rigide di lunghezza  $l$  ed  $l'$ , composte da cilindri rigidi aventi diametro  $D$  e lunghezza  $X_0 D$ , il contributo del volume escluso può essere scritto come:

$$v_{excl}(l, l') = 2 \int f(\mathbf{u}) f(\mathbf{u}') D^3 \left[ \frac{\pi}{2} \sin \gamma + \frac{\pi}{2} X_0 (1 + |\cos \gamma|) + \frac{4}{\pi} E(\sin \gamma) \frac{l+l'}{2} + 2(\sin \gamma) X_0^2 l l' \right] d\Omega d\Omega' \quad (11.23)$$

dove  $\cos \gamma = \mathbf{u} \cdot \mathbf{u}'$  ed  $E(\sin \gamma)$  l'integrale ellittico definito:

$$E(\sin \gamma) = \frac{1}{4} \int_0^{2\pi} \sqrt{1 - \sin^2 \gamma \sin^2 \Psi} d\Psi. \quad (11.24)$$

Queste espressioni appena discusse sono valide nel caso di cilindri rigidi (HC) ma sono state supposte valide anche nel caso della simulazione di particelle descritte mediante superquadriche (SQ). Ovviamente nell'ipotesi di avere dei cilindri rigidi piegati esse varranno esclusivamente nel caso in cui il "bending" sia nullo, cioè in prima approssimazione valide nel caso in cui  $\Theta \rightarrow 180^\circ$ , mentre ad angolature differenti sarà necessario rivalutare la formulazione dell'entropia orientazionale  $\sigma_0$  a causa della rottura della simmetria di rotazione azimutale. Inoltre per colpa della differente geometria, anche il calcolo del volume escluso nel nostro caso sarà differente pur supponendo ancora valida l'assunzione che esso scali con le lunghezze  $l$  ed  $l'$ .

In realtà anche questa seconda affermazione potrebbe non essere del tutto veritiera, Infatti i modi che abbiamo per disporre due HBC consecutivamente sono infiniti a causa della perdita di simmetria rotazionale attorno all'asse del cilindro. Questo fatto influenzerà certamente

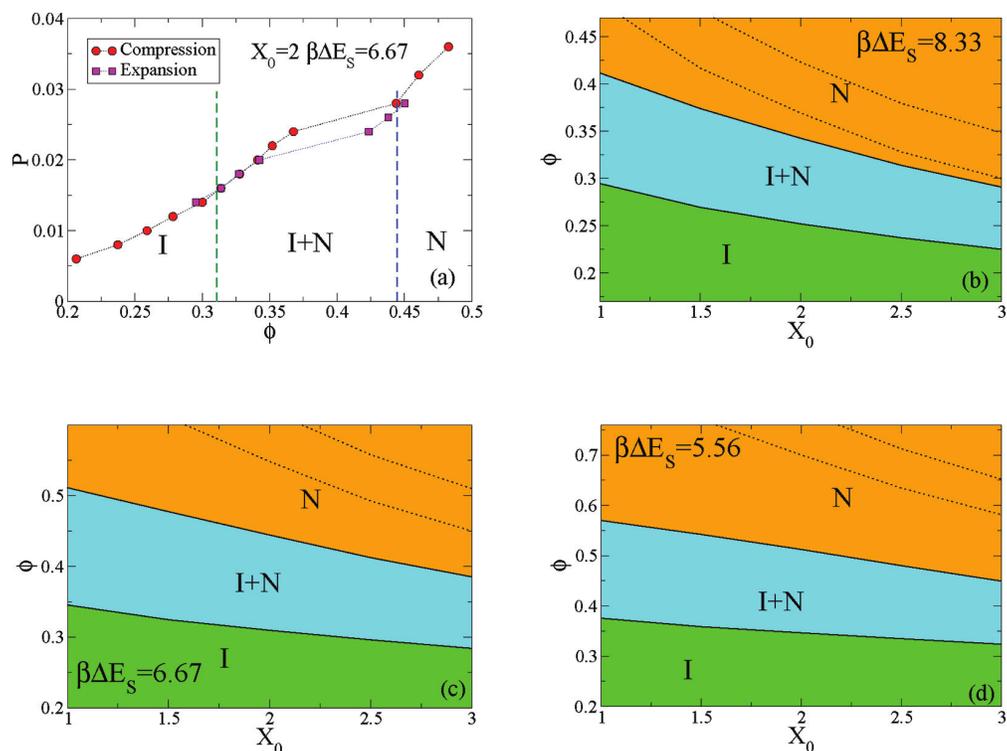


Figura 11.1: In alto a destra viene rappresentato il diagramma di fase ottenuto da simulazioni MC di superquadriche con  $X_0 = 2$  con  $\beta\Delta E_S = 6.67$ . Gli altri diagrammi rappresentano le regioni di coesistenza delle fasi nematiche ed isotropiche teoriche al variare degli “aspect-ratio” e delle energie di “stacking”. L’immagine è stata presa dall’articolo di De Michele [19].

anche il contributo entropico che però, sempre nel limite  $\Theta \rightarrow 180^\circ$ , supponiamo essere pari a:

$$\sigma_0(l) = \frac{1}{8l_p} \int \left( \frac{\partial f}{\partial \theta} \right)^2 f^{-1} d\Omega - 2 \ln \left[ \int f^{1/2} d\Omega \right] + \ln(4\pi) \quad \text{se } l_p \gg l \quad (11.25)$$

$$\sigma_0(l) = \int f \ln(4\pi f) d\Omega + \frac{1}{12l_p} \int \left( \frac{\partial f}{\partial \theta} \right)^2 f^{-1} d\Omega \quad \text{se } l_p \ll l \quad (11.26)$$

secondo quanto calcolato da Khokhlov e Semenov [58, 59].

Alla luce delle considerazioni fatte finora, l’energia libera può essere riscritta in questi termini:

$$\begin{aligned}
\frac{\beta F}{V} &= \frac{2}{3l_p} \sum_{l=1}^{\infty} l \int [f^{1/2}] \nabla^2 [f^{1/2}] d\Omega + \sum_{l=1}^{\infty} \int f_l(\mathbf{u}) \{\ln [4\pi v_d f_l(\mathbf{u})] - 1\} - \\
&+ \frac{\eta(\varphi)}{2} \sum_{l,l'=1}^{\infty} \int f_l(\mathbf{u}) f_{l'}(\mathbf{u}') v_{excl}(l, l') d\Omega d\Omega' + \\
&- (\beta \Delta E_S + \sigma_b) \sum_{l=1}^{\infty} \int (l-1) f_l(\mathbf{u}) d\Omega
\end{aligned} \tag{11.27}$$

formulazione usata anche nel lavoro di De Michele [19]. Come già accennato sappiamo che queste approssimazioni non valgono esattamente nel caso di HBC da noi trattato, ma le supporremo valide per confrontarci con tutti gli articoli citati sinora, soprattutto per paragonare i risultati che otterremo. Se la nostra idea dell'introduzione del "bending" ci porterà ai risultati aspettati, sarà interessante nel prossimo futuro aggiornare la teoria che fa da sfondo alle nostre simulazioni Monte Carlo, in particolare prestando particolare attenzione riguardo al calcolo del contributo entropico orientazionale presente nella formulazione dell'energia libera (11.27).

### 11.3.1 Fase isotropica

Quando il sistema si trova in una fase isotropica tutte le orientazioni possibili degli oggetti all'interno del box sono equiprobabili, perciò avremo che la probabilità di orientazione del cilindro sarà:

$$f(\mathbf{u}) = \frac{1}{4\pi}. \tag{11.28}$$

Inserendolo nell'equazione (11.20) e calcolando gli integrali su tutte le direzioni possibili otteniamo che l'energia libera si può ridurre alla forma:

$$\begin{aligned}
\frac{\beta F}{V} &= \sum_{l=1}^{\infty} \nu(l) \{\ln [v_d \nu(l)] - 1\} + \frac{\eta(\varphi)}{2} \sum_{l,l'=1}^{\infty} \nu(l) \nu(l') v_{excl}(l, l') + \\
&- (\beta \Delta E_S + \sigma_b) \sum_{l=1}^{\infty} (l-1) \nu(l)
\end{aligned} \tag{11.29}$$

mentre il volume escluso può essere calcolato esplicitamente:

$$v_{excl}(l, l') = \frac{\pi^2}{8} D^3 + \left( \frac{3\pi}{8} + \frac{\pi^2}{8} \right) [l + l'] X_0 D^3 + \frac{\pi}{2} ll' X_0^2 D^3. \tag{11.30}$$

Inoltre è possibile assumere che la distribuzione della lunghezza degli aggregati lineari sia

descritta da una funzione esponenziale con valore medio  $M$ :

$$\begin{aligned}\nu(l) &= \rho M^{-(l+1)} (M-1)^{l-1} \\ &= \frac{\rho}{M(M-1)} e^{-l[\ln M - \ln(M-1)]}\end{aligned}\quad (11.31)$$

con  $M$  lunghezza media dell'aggregato ottenuta come

$$M = \frac{\sum_1^\infty l \nu(l)}{\sum_1^\infty \nu(l)} \quad (11.32)$$

che ci permette di riscrivere l'energia libera in termini di questa grandezza:

$$\begin{aligned}\frac{\beta F}{V} &= -\rho(\beta \Delta E_S + \sigma_b)(1 - M^{-1}) + \eta(\varphi) \left[ B_I X_0^2 + \frac{v_d k_I}{M} + \frac{A_I}{M^2} \right] \rho^2 + \\ &+ \frac{\rho}{M} \left[ \ln \left( \frac{v_d \rho}{M} \right) - 1 \right] + \rho \frac{M-1}{M} \ln(M-1) - \rho \ln M\end{aligned}\quad (11.33)$$

in cui i parametri  $A_I$ ,  $B_I$  e  $k_I$  dipendono in generale da  $X_0$  ed ottenibili mediante confronto con l'equazione polinomiale per il volume escluso:

$$v_{excl}(l, l') = 2 \left[ A_I(X_0) + k_I(X_0) v_d \frac{l+l'}{2} + B_I(X_0) X_0^2 l l' \right]. \quad (11.34)$$

Dalla minimizzazione dell'energia libera rispetto a  $M$  e scartando i termini di secondo grado otteniamo infine:

$$M = \frac{1}{2} \left( 1 + \sqrt{1 + 4\omega \varphi e^{k_I \varphi \eta(\varphi) + \beta \Delta E_S}} \right) \quad (11.35)$$

avendo definito  $\omega = e^{\sigma_b}$ .

### 11.3.2 Fase nematica

Per quanto concerne la fase nematica del sistema invece, la funzione  $f(u)$  dipende da un set di parametri ottenibili generalmente attraverso la minimizzazione dell'energia libera. Anche in questo caso possiamo assumere una distribuzione esponenziale per la  $\nu(l)$  mentre per la distribuzione angolare sfruttiamo la forma proposta da Onsager [14]:

$$f(\mathbf{u}) = f_O(\mathbf{u}) = \frac{\alpha}{4\pi \sinh \alpha} \cosh(\alpha \cos \theta) \quad (11.36)$$

dove con  $\theta$  abbiamo indicato l'angolo formato tra  $\mathbf{u}$  e l'asse di nematizzazione.

Analogamente a quanto fatto in precedenza per il caso isotropico scegliamo, per il calcolo

del volume escluso di due DNA mediato sull'angolo solido, l'espressione:

$$v_{excl}(l, l', \alpha) = 2 \left[ A_N(\alpha) + k_N(\alpha) v_d \frac{l+l'}{2} + B_N(\alpha) X_0^2 l l' \right] \quad (11.37)$$

che, inserita assieme alla 11.31 nella 11.20, ci permette di ricavare l'espressione dell'energia libera per un sistema nella fase nematica:

$$\begin{aligned} \frac{\beta F}{V} = & \hat{\sigma}_0 - \rho(\beta \Delta E_S + \sigma_b)(1 - M^{-1}) + \eta(\varphi) \left[ B_N(\alpha) X_0^2 + \frac{v_d k_N(\alpha)}{M} + \frac{A_N(\alpha)}{M^2} \right] \rho^2 + \\ & + \frac{\rho}{M} \left[ \ln \left( \frac{v_d \rho}{M} \right) - 1 \right] + \rho \frac{M-1}{M} \ln(M-1) - \rho \ln M \end{aligned} \quad (11.38)$$

avendo definito  $\hat{\sigma}_0 = \sum_l \sigma_0(l) \nu(l)$ .

### 11.3.3 Regione di coesistenza

Usando i funzionali di energia libera calcolati nella fase isotropica (11.33) e nematica (11.38) possiamo calcolare gli estremi della regione di coesistenza  $\varphi_I - \varphi_N$  minimizzando le espressioni dell'energia libera rispetto a la lunghezza media delle catene nella fase isotropica e nematica, cioè  $M_I$  ed  $M_N$ , e rispetto al parametro  $\alpha$ .

Inoltre dobbiamo richiedere che la regione di coesistenza rispetti i vincoli di pressione,  $P_I = P_N$ , e del potenziale chimico,  $\mu_I = \mu_N$ , che caratterizzano la transizione di fase del primo ordine.

In questo modo dobbiamo risolvere numericamente le seguenti espressioni:

$$\frac{\partial}{\partial M_I} F_I(\rho_I, M_I) = 0 \quad (11.39)$$

$$\frac{\partial}{\partial M_N} F_N(\rho_N, M_N, \alpha) = 0 \quad (11.40)$$

$$\frac{\partial}{\partial \alpha} F_N(\rho_N, M_N, \alpha) = 0 \quad (11.41)$$

$$P_I(\rho_I, M_I) = P_N(\rho_N, M_N, \alpha) \quad (11.42)$$

$$\mu_I(\rho_I, M_I) = \mu_N(\rho_N, M_N, \alpha). \quad (11.43)$$

Come abbiamo già detto non calcoleremo esplicitamente questi parametri, anche perché la teoria andrebbe riadeguata al caso di oggetti che non conservano più la degenerazione azimutale di rotazione attorno all'asse  $\mathbf{u}$  a causa del piegamento dello stesso.

Quindi ci riferiremo ai valori calcolati in [20, 19] come riferimento ad una struttura a doppio cilindro a "bending" nullo, valori a cui dovranno tendere i risultati ottenuti nel caso di cilindri piegati presi in esame in questa sede.

## Capitolo 12

# Simulazioni MC di Cilindri Piegati

### 12.1 Modellizzazione di DNA piegati

Avevamo già accennato a come modellizzare il fatto che il DNA piegasse su se stesso, occorre ora esaminare più nel dettaglio la situazione. Poiché le sequenze considerate in questo contesto sono palindrome e la disposizione della sequenza è simmetrica rispetto al baricentro della molecola stessa, possiamo considerare la metà costituita dalle prime sei coppie come del tutto equivalente alla seconda metà restante.

Supponendo che la molecola si pieghi rispetto ad un asse passante per il baricentro è lecito quindi approssimare ciascuna delle due sotto unità con una rappresentazione “coarse-grained” cilindrica. Abbiamo così ottenuto due cilindri, piegati l’uno rispetto all’altro ed i cui assi ed i centri delle basi interne passano per il baricentro della molecola.

Poiché stiamo per simulare tali oggetti mediante dinamica MC al fine di ottenere le proprietà termodinamiche del sistema è necessario evitare eventuali “incastrì” non realistici dei doppi cilindri, dovuti all’apertura di una fessura in prossimità del centro dell’oggetto come precedentemente schematizzato in figura 8.5. Per fare questo prolunghiamo l’altezza del cilindro di ciascuna delle due basi centrali in modo tale da riempire tale fessura e scongiurare eventuali situazioni sfavorevoli entropicamente.

L’angolo di “bending” che scegliamo sarà scelto contestualmente ai valori degli angoli ricavati mediante la media delle osservabili che avevamo studiato nella seconda parte del testo, pur tenendo conto del fatto che non siamo riusciti a stimarlo facendo uno studio completo della sequenza a causa degli effetti legati al “fraying” delle ultime coppie di basi. Faremo quindi uno studio in funzione dell’angolo, provando a simulare diversi gradi di piegatura e cercando di sondare il più vasto range di angoli compatibilmente con quanto osservato nelle simulazioni a dinamica molecolare. In particolare ci spingeremo a considerare oggetti con angoli che variano con valori compresi tra  $\Theta = 134^\circ \div 160^\circ$ .

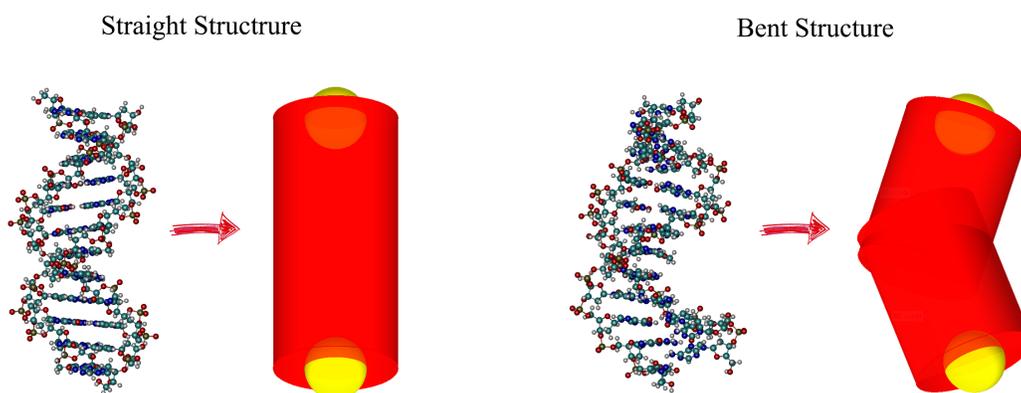


Figura 12.1: In questa immagine è mostrata la modellizzazione “coarse-grained” del DNA mediante l’utilizzo di cilindri e basi attrattive. A sinistra notiamo come la configurazione cristallografica dell’acido nucleico sia ben approssimabile con un cilindro rettilineo ma a causa della sua deformazione strutturale in soluzione la rappresentazione HBC di destra è da preferire nelle simulazioni Monte Carlo.

A causa del prolungamento introdotto per evitare effetti di incastrici configurazionali vari l’“aspect ratio” di tale sistema, considerando il doppio cilindro in configurazione rettilinea, è pari a  $X_0 = 2.15$ , valore prossimo a quello calcolato per la situazione iniziale di  $X_0 = 1.75$ .

Analogamente a quanto visto nell’articolo di riferimento andremo a modellizzare le interazioni tra siti attrattivi posti su ciascuna delle due basi estremali, attraverso un potenziale di tipo “square-well” (SW):

$$\beta U_{sw} = \begin{cases} -\beta \Delta E_S & \text{se } r < \delta \\ 0 & \text{se } r > \delta \end{cases} \quad (12.1)$$

con  $\delta = 1.22$ , distanza alla quale viene troncata l’interazione dal punto a distanza  $X_0 - 0.46$  dal centro di massa del singolo cilindro. Tale scelta è allineata a quanto fatto da De Michele [19] per riprodurre efficacemente il range di azione delle forze di “stacking” idrofobico, anche se in questo contesto prenderemo in esame solamente il valore  $\beta \Delta E_S = 8.33$  valore che, da quanto emerso da tale articolo, sembrerebbe sovrastimare l’energia di “stacking”. Decidiamo di optare per tale valore poiché utilizzando un’energia abbastanza grande osserveremo la transizione di fase a valori inferiori della “volume fraction”, consentendoci di esplorare un range abbastanza vasto di valori prima di raggiungere densità paragonabili alla frazione di impacchettamento dei cilindri, oltre la quale non è più possibile comprimere il sistema.

## 12.2 Volume degli oggetti

Poiché ci stiamo proponendo di studiare un sistema formato da oggetti costituiti da due cilindri intersecati con assi non paralleli tra loro, il calcolo del volume di ogni HBC non è propriamente un'operazione banale da eseguire.

Piuttosto che calcolare esplicitamente l'espressione per il volume degli stessi decidiamo di utilizzare ancora una volta l'algoritmo Monte Carlo per ottenere una stima degli stessi. Attraverso un programma scritto appositamente in C definiamo un box cubico di volume  $V_{box}$  all'interno del quale è contenuto il nostro doppio cilindro piegato. Generiamo  $10^6$  punti casualmente all'interno del box e valutiamo quanti di essi cadono all'interno dell'HBC rispetto al totale considerato. Grazie a questo è facile stimare il volume dell'oggetto:

$$V_{HBC} = V_{box} \frac{(\# \text{punti interni HBC})}{10^6} \quad (12.2)$$

ottenendo come risultato finale i seguenti valori:

$$V_{HBC}^{134^\circ} = 14.458 \quad V_{HBC}^{138^\circ} = 14.291 \quad V_{HBC}^{141^\circ} = 14.167 \quad (12.3)$$

$$V_{HBC}^{150^\circ} = 13.871 \quad V_{HBC}^{155^\circ} = 13.750 \quad V_{HBC}^{160^\circ} = 13.656 \quad (12.4)$$

la cui distribuzione è riportata nel grafico in figura 12.2.

Il volume del doppio cilindro in assenza di piegatura risulta essere  $V_{HBC}^{180^\circ} = 13.511$  pari cioè al volume di un singolo cilindro con diametro  $D = 2$  ed "aspect-ratio"  $X_0 = 2.15$ :

$$V = \frac{1}{4} X_0 D^3 \pi = \frac{1}{4} L D^2 \pi = 13.509 \quad (12.5)$$

in accordo con quanto ottenuto dal calcolo MC del volume.

## 12.3 Confronto con i dati sperimentali

Come abbiamo appena visto il volume degli oggetti HBC dipende in modo sensibile dall'angolazione tra gli assi di ciascun cilindro a causa degli effetti legati alla sovrapposizione dei due cilindri.

Tenendo conto di questo fatto sarà utile, per confrontarci con le simulazioni Monte Carlo di De Michele [19], introdurre la grandezza chiamata "volume fraction" che rappresenta il rapporto tra il volume effettivo degli oggetti contenuti nel sistema  $N \cdot V_{HBC}$  rispetto al volume totale  $V$  a loro disposizione:

$$\varphi = \frac{N \cdot V_{HBC}}{V}. \quad (12.6)$$

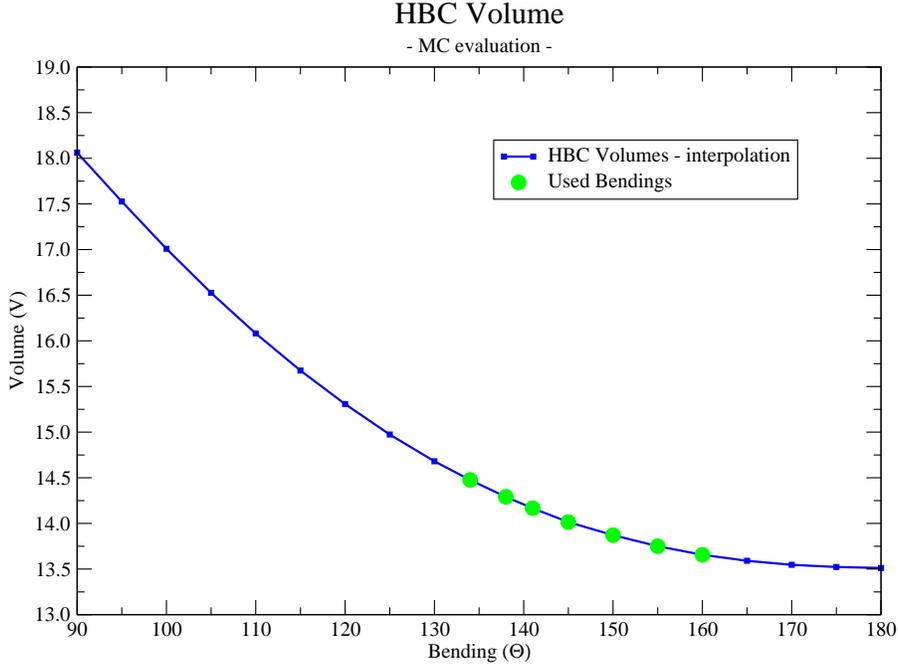


Figura 12.2: Stime del volume degli oggetti HBC ottenuti mediante calcolo Monte Carlo per tutti gli angoli da  $90^\circ$  al  $180^\circ$ . Notiamo che all’aumentare del “bending”, cioè quando diminuisce  $\Theta$ , il volume dei singoli HBC aumenta a causa del diminuire degli effetti di sovrapposizione tra i singoli cilindri. In verde sono riportati i valori per gli angoli utilizzati in questa tesi.

Grazie a questa quantità potremo graficare il diagramma di fase dei sistemi HBC studiati, ricavando quindi le “volume fraction” di nematizzazione, ovvero le frazioni di impacchettamento critiche a cui il sistema subisce la transizione di fase. Chiamando  $\varphi_I$  l’ultimo valore dopo il quale il sistema inizia la transizione di fase e  $\varphi_N$  il primo valore della frazione di volume a cui il sistema nematizza, potremo confrontare questi valori con quanto misurato sperimentalmente da Zanchetta et al. [16] nei casi da loro presi in esame.

Occorre quindi trovare un fattore che ci consenta di legare  $\varphi$  alla concentrazione critica a cui il sistema transisce ma poiché sappiamo che la densità è legata alla concentrazione  $c$  mediante la formula:

$$\rho = \frac{c}{m_{DNA}} \quad (12.7)$$

possiamo convertire la concentrazione di DNA all’interno del sistema con la “volume fraction” mediante:

$$\varphi = \rho V = \frac{LD^2\pi}{4m_{DNA}} c. \quad (12.8)$$

oppure equivalentemente con la:

$$\varphi = \frac{0.172D^3\pi}{8m_N}c \quad (12.9)$$

dove  $m_N$  è la massa del singolo nucleotide.

Esplicitando la concentrazione critica in funzione dell'angolo di "bending"  $c_N(\Theta)$  a cui il sistema transisce nello stato nematico  $\varphi_N$ , otteniamo infine che:

$$c_N(\Theta) = \varphi_N(\Theta) \frac{4m_{DNA}}{LD^2\pi} \quad (12.10)$$

che espressa nelle opportune unità  $mg/ml$  ci consentirà di confrontare quanto ottenuto con i dati sperimentali.

In realtà poiché abbiamo calcolato esplicitamente il volume degli oggetti HBC possiamo effettuare una conversione più accurata con quanto ottenuto dalle simulazioni MC. Poiché sappiamo che il volume del singolo cilindro rettilineo è pari a  $V = \frac{LD^2}{4}\pi$  possiamo riscrivere la relazione che lega la volume fraction alla concentrazione del DNA nel sistema come:

$$c(\Theta) = \frac{\varphi(\Theta)m_{DNA}}{V} \quad \varphi(\Theta) = \frac{V}{m_{DNA}}c(\Theta). \quad (12.11)$$

## 12.4 Definizione delle osservabili di nematizzazione

Per poter apprezzare la transizione nematica all'aumentare della frazione di impacchettamento del sistema, è necessario introdurre una grandezza che descriva il grado di ordinamento nematico degli oggetti considerati.

Nel caso già studiato di cilindri rigidi e superquadriche, è ragionevole pensare che il vettore  $\mathbf{u}$  che indichi la direzione di nematizzazione sia rappresentato dalla media dei vettori che congiungono i centri delle basi degli stessi. Esso giace quindi esattamente sull'asse molecolare. Poiché l'introduzione del grado di piegatura dei cilindri determina la rottura di simmetria assiale, non sarà più possibile usare questo parametro nel considerare il caso di doppi cilindri piegati.

E' però ragionevole supporre che il nuovo vettore  $\mathbf{u}_{HBC}$  che descrive l'asse di nematizzazione tenda al vettore assiale del cilindro rettilineo quando  $\Theta = 180^\circ$ , cioè:

$$\mathbf{u}_{HBC} \xrightarrow{\Theta \rightarrow 180^\circ} \mathbf{u}. \quad (12.12)$$

In questo modo abbiamo la possibilità di esprimere il grado di ordinamento nematico del sistema, ancora una volta come fatto in [19], mediante il calcolo dell'autovalore più grande  $S$

del tensore di Landau-De Gennes  $Q$  definito come [60]:

$$Q_{\alpha\beta} = \frac{1}{N} \sum_i \frac{3}{2} \left\langle (\mathbf{u}_i^{HBC})_\alpha (\mathbf{u}_i^{HBC})_\beta \right\rangle - \frac{1}{2} \delta_{\alpha\beta} \quad (12.13)$$

dove  $\alpha, \beta \in \{x, y, z\}$  mentre  $(\mathbf{u}_i^{HBC})_\alpha$  rappresenta la componente  $\alpha$ -esima del vettore unitario definito sulla congiungente dei due centri delle basi estremali del doppi cilindro.

Come mostrato chiaramente in figura 12.3, dove si evidenzia anche il vettore  $\mathbf{u}_{HBC}$  attraverso la congiungente dei due centri (in giallo) delle basi estremali, sono possibili numerose tipologie di configurazione tra gli HBC e prendendo i due casi estremi abbiamo che:

- Gli oggetti 1 – 2 – 3 – 4 sono disposti su una geometria circolare, essa rappresenta il caso peggiore di concatenamento ed ovviamente non corrisponde ad una fase nematica;
- Gli oggetti 4 – 5 – 6 – 7 invece sono effettivamente concatenati ed allineati grossomodo sullo stesso asse principale, è quindi questa la situazione che ci aspetteremo durante la nematizzazione del sistema.

A causa del grandissimo numero di configurazioni possibili per il sistema, vedremo che il raggiungimento della completa nematizzazione sarà un processo estremamente lungo e difficile da raggiungere, necessitando di un gran numero di step MC da eseguire durante la simulazione. La figura inoltre schematizza una semplice situazione bidimensionale mentre la disposizione tridimensionale di tali oggetti può essere ancora più complessa e formare aggregati molto contorti.

Per come è stato definito,  $S$  potrà assumere valori compresi tra  $S \in [0; 1]$  dove il valore  $S = 1$  corrisponde al caso di perfetto allineamento assiale mentre il caso  $S = 0$  rappresenta la fase completamente isotropica del sistema. Ai fini di questa tesi considereremo il sistema come completamente transitato nella fase nematica, quando il valore dell'autovettore  $S_{nem}$  supera la soglia  $S > 0.5$  stabilizzandosi attorno ad un valore di equilibrio.

## 12.5 Set-up sperimentale del sistema

Per allinearci a tutti i valori angolari trovati per ogni tipo di sequenza di DNA nel capitolo precedente, decidiamo di fare uno studio del diagramma di fase per sette valori di angoli  $\Theta$  di “bending” dei doppi cilindri piegati, scelti contestualmente a quanto ottenuto nella seconda parte della tesi dalle osservabili fosforiche e di Dickerson.

Scegliamo, contestualmente a quanto fatto nel lavoro [19], di metterci in una situazione tale per cui il contributo dell'energia di “stacking” tra basi e temperatura del sistema sia tale che  $\beta\Delta E_S = 8.33$ . La temperatura in questo caso è settata al valore fissato  $T = 0.12$ , poiché con valori troppo bassi potremmo incorrere in effetti legati all'eccessiva lunghezza della vita

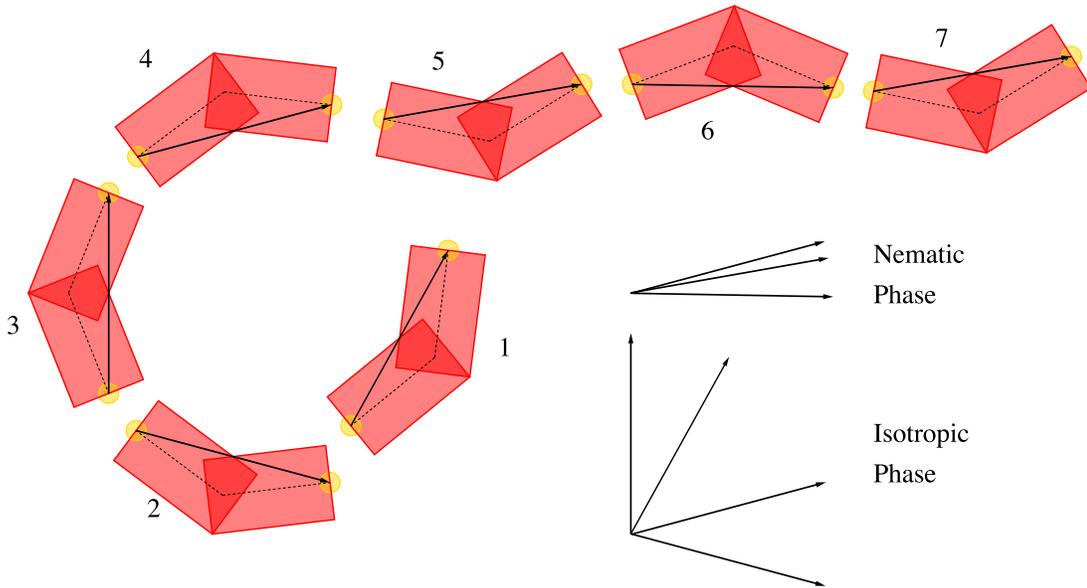


Figura 12.3: Schematizzazione bidimensionale delle possibili disposizioni di oggetti HBC. Solo gli oggetti 4 – 5 – 6 – 7 sono considerati disposti in fase nematica, gli altri rientrano in configurazioni a carattere isotropico.

media dei legami, che limiterebbero una corretta esplorazione dello spazio delle fasi del sistema mentre con temperature più alte potremmo incappare in fasi colonnari o cristalline perdendoci quindi la fase nematica a cui siamo interessati.

Da quanto emerso nella pratica, partendo con una disposizione completamente casuale degli oggetti all'interno del box di simulazione, abbiamo riscontrato una grande difficoltà da parte del sistema a raggiungere una fase nematica a causa delle molte mosse possibili ai singoli HBC specialmente per i gradi di piegatura fino a  $\Theta = 141^\circ$ .

Per colpa di questo fatto occorrerebbe effettuare un numero enorme di mosse MC per raggiungere il valore aspettato per la "volume fraction" del sistema, ma non avendo a disposizione molto tempo decidiamo di usare alcuni accorgimenti per facilitare il raggiungimento della fase nematica. Per prima cosa creiamo un box a forma di parallelepipedo a due facce quadrate di lato  $l_{box} = 20.5$  e lato lungo  $L_{box} = 60.0$ , il cui volume risulta quindi avere valori di  $V_{box}^{sim} \simeq 25000$ . In questo modo, a causa della maggior lunghezza possibile nella formazione di catene, l'asse nematico favorito sarà circa coincidente con una delle diagonali passanti per il centro del box. All'interno di tale box disponiamo i doppi cilindri già allineati tra loro, a formare una specie di fase nematica iniziale con asse parallelo al lato lungo del box: Questa però non è effettivamente una fase nematica stabile per il sistema pur partendo con gli  $\mathbf{u}_{HBC}$  già tutti paralleli, infatti lungo la catena la perfetta nematizzazione prevede una rotazione di  $180^\circ$  attorno all'asse che congiunge i centri delle basi, perciò non stiamo partendo

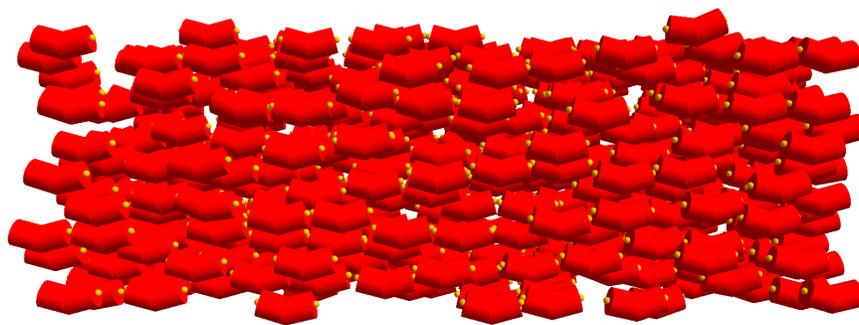


Figura 12.5: Configurazione nematica non equilibrata, preparata per l’inizializzazione di tutti i sistemi a “bending” differente. In questo caso è riportata la situazione iniziale nel caso di  $150^\circ$ , del tutto analoga a quella delle altre.

da una configurazione vantaggiosa dal punto di vista entropico per la transizione di fase. Dopo pochi passi MC infatti (all’incirca dopo  $2 \cdot 10^4$  step), la fase nematica instabile iniziale viene persa completamente ed il sistema rapidamente si porta in configurazioni isotropiche pur raggiungendo più facilmente la nematizzazione a determinate pressioni critiche.

Il numero di oggetti scelti per la configurazione viene settato a 500, ritenuto un giusto compromesso tra velocità di computazione ed efficienza termodinamica, mentre per le mosse MC da effettuare eseguiamo tre run consecutivi da  $2 \cdot 10^6$  step prendendo in esame solo l’ultimo.

Il primo gruppo di simulazioni verrà effettuato comprimendo il sistema, partendo cioè da volumi maggiori ed aumentando la pressione a step di  $\Delta P = 0.001$ . Successivamente, osservando il diagramma di fase e l’andamento del parametro d’ordine, scegliamo il valore di pressione a cui siamo convinti che il sistema abbia nematizzato. Prendiamo l’ultima configurazione ottenuta dal programma di simulazione ed usiamola come base di partenza per ulteriori simulazioni, stavolta comprimendo e decomprimendo il sistema.

Questo è necessario perché, essendo i primi run di compressione indipendenti tra loro, ogni

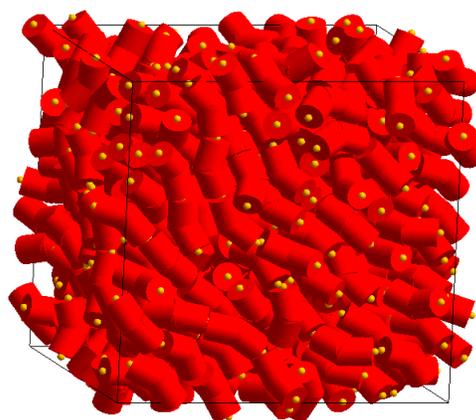


Figura 12.4: Osservazione dell’avvenuta nematizzazione del sistema formato da  $\Theta = 141^\circ$  partendo da oggetti casualmente distribuiti all’interno di un volume cubico.

punto della fase nematica possiederà assi nematici differenti, rappresentando situazioni termodinamiche differenti: scegliendo una configurazione nematica e procedendo in compressione e decompressione saremo sicuri di percorrere la stessa isoterma mantenendo l'asse nematico fisso e preservando la direzione in cui era avvenuto l'ordinamento.

## Capitolo 13

# Analisi dati

Una volta concluse le simulazioni Monte Carlo di doppi cilindri piegati è possibile passare alla fase di analisi dei dati ottenuti. In questo caso la situazione è differente rispetto a quanto fatto con la simulazione di dinamica molecolare in cui, una volta completata la simulazione, potevamo prendere direttamente le traiettorie e studiarne l'evoluzione temporale. Per le simulazioni MC occorrerà effettuare un'analisi statistica delle grandezze definite precedentemente, per poter mediare sulle fluttuazioni ed ottenere i valori medi delle osservabili solo dopo aver osservato graficamente la stabilizzazione del sistema.

In questo caso, infatti, non è possibile stabilire un numero minimo di step MC dopo il quale tutti i sistemi abbiano stabilizzato effettuando l'analisi dati da questo punto in poi, per colpa della grande dipendenza dalla pressione  $P$  che stiamo esaminando e soprattutto dal bending delle strutture HBC considerate.

Proprio per questo motivo cerchiamo di agire in questo modo. Come già accennato abbiamo effettuato due run da  $2 \cdot 10^6$  passi MC consecutivi per ciascuna sequenza partendo da una configurazione allineata casuale dei doppi cilindri e comprimendo il sistema aumentando la pressione. Il primo dei due run è stato scartato, considerandolo un run di equilibratura del sistema nel quale esso riesce a raggiungere una fase stabile, mentre sul secondo è stata fatta l'analisi degli ultimi  $10^6$  dati ottenuti. Dai valori delle "volume fraction" medie ricavate in questo modo abbiamo stimato, sia attraverso l'osservazione del diagramma di fase, sia attraverso l'analisi dell'andamento del parametro d'ordine  $S$  della transizione, il punto in cui avveniva la transizione di fase registrandone i valori di pressione e volume ottenuti. Ovviamente questo valore sovrastima il punto esatto a cui il sistema nematizza, infatti durante la fase di compressione il sistema mantiene una certa metastabilità nella fase isotropica prima di effettuare la transizione di fase, come mostrato anche in figura 13.11 nel caso di simulazione con superquadriche.

In questo modo abbiamo ottenuto i diagrammi di fase dei sistemi, anche se in realtà ciascuno dei punti del diagramma è indipendente dagli altri poiché ottenuto partendo da una

stessa situazione iniziale e comprimendo. A causa di eventuali effetti di taglia finita è possibile che i punti in cui il sistema ha transito nella fase nematica rappresentino situazioni con asse nematico differente che porterebbe la lunghezza media dell'aggregato a risentire della direzione di nematizzazione all'interno del box. Per questo motivo è ragionevole supporre che nello stato isotropico questo fenomeno sia irrilevante mentre sono da valutare con attenzione i risultati ottenuti a pressioni in cui il sistema risulta transito nello stato nematico.

Per fare in modo che l'isoterma rappresenti una stessa situazione termodinamica indipendentemente dalla direzione dell'asse di nematizzazione decidiamo di agire in questo modo. Comprimiamo il sistema finché non riscontriamo la discontinuità del parametro d'ordine, che è un indice dell'avvenuta nematizzazione del sistema e successivamente constatiamo che per i punti successivi il sistema si sia grossomodo stabilizzato in una configurazione liquido-cristallina, osservando che il parametro d'ordine stabilizza a valori superiori a  $S = 0.5$ . A questo punto prendiamo la prima configurazione nematica a cui riteniamo completa la stabilizzazione scartando i risultati a pressioni superiori a quella ottenuta. In questo modo abbiamo un sistema formato da HBC transito completamente in uno stato nematico e con asse di nematizzazione ben definito, da cui possiamo partire per esaminare ancora una volta l'isoterma stavolta effettuando dei run in decompressione. Anche in questo caso noteremo una certa isteresi nella transizione nematica-isotropica che sottostima la densità a cui avviene esattamente la transizione ma che unita alla metastabilità ottenuta durante la fase di compressione ci consentirà di valutare la "volume fraction" a cui il sistema cambia il proprio stato di aggregazione. In realtà partendo sempre dalla prima configurazione nematica ottenuta comprimeremo ancora il sistema per osservare l'eventuale presenza di altre transizioni (come per esempio la transizione nematica-colonnare) e completare il diagramma di fase.

Grazie a queste analisi riusciremo a stimare le concentrazioni critiche a cui i sistemi cambiano fase al variare del grado di "bending" che caratterizza i singoli costituenti del sistema. In particolare suddivideremo la nostra analisi in due gruppi di sistemi HBC: il primo con piegature  $\Theta > 150^\circ$  in cui abbiamo riscontrato una nematizzazione completa ed il secondo, con  $\Theta < 150^\circ$ , per cui la stabilizzazione richiede un numero maggiore di step MC di quanto esaminato in questa tesi.

### 13.1 Equilibratura del sistema

Come abbiamo detto è difficile stabilire un numero minimo di step MC che, partendo dalla medesima disposizione degli HBC, ci consenta di ottenere dei valori ben rappresentativi per le medie del sistema per ogni grado di piegatura trattato. Cerchiamo perciò delle linee guida fondamentali per poter intraprendere un'analisi che sia la più accurata possibile dei dati a nostra disposizione. Ovviamente avendo a disposizione delle simulazioni con un numero ancora

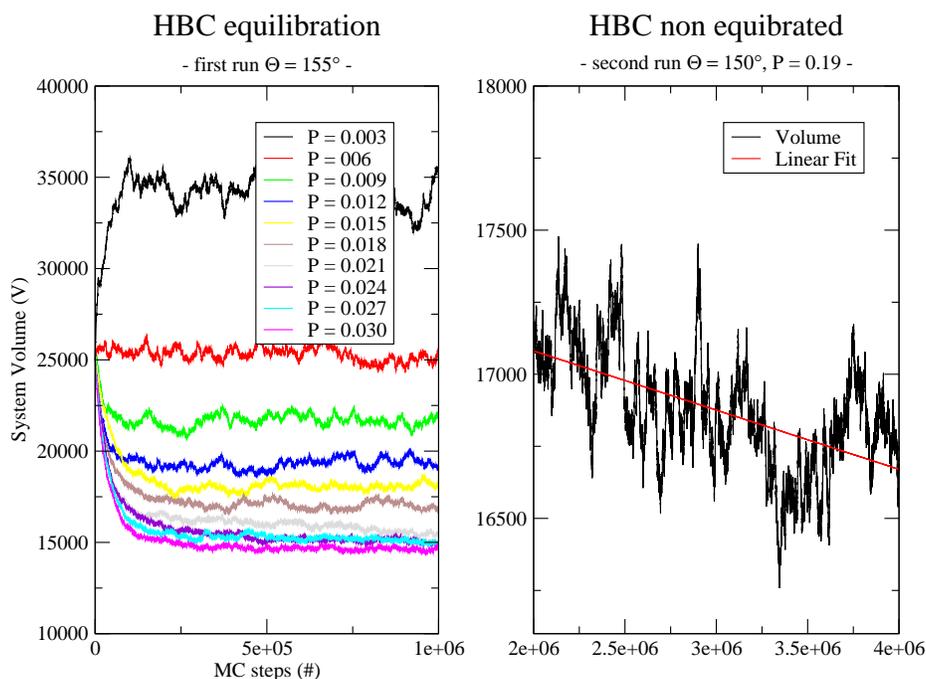


Figura 13.1: Stabilizzazione del volume per un sistema composto da HBC con  $\Theta = 155^\circ$ . A sinistra mostriamo come ad angolature inferiori  $\Theta = 150^\circ$  il sistema nematico necessita di ulteriori step MC per completare la sua equilibratura rispetto ai  $4 \cdot 10^6$  che siamo riusciti a simulare in questa sede.

maggiore, ad esempio  $10^7 \div 10^8$ , di step MC eseguiti tale problema non sopraggiungerebbe e potremmo effettuare un'analisi più fine dei diagrammi di fase di ciascun sistema.

Come abbiamo già detto, ogni run parte da una configurazione nematica non equilibrata e viene compresso a determinate pressioni, con un intervallo di  $\Delta P = 0.001$  per un primo gruppo di angoli mentre  $\Delta P = 0.003$  per il secondo. Operando in questo modo il sistema avrà bisogno di un certo numero di step MC prima di portarsi al suo volume caratteristico e dal comportamento del primo set run da  $2 \cdot 10^6$  possiamo cercare di stabilire quando il volume comincia a fluttuare attorno ad un valore stabile.

Ad una prima analisi del grafico in figura 13.1, nel quale viene riportato il primo run di un sistema HBC a  $\Theta = 155^\circ$ , notiamo che ad ogni pressione il sistema si porta rapidamente in una configurazione stabile attorno alla quale iniziano delle fluttuazioni dell'ordine dell'1%. Verrebbe da prendere, quindi, il valore di  $2.50 \cdot 10^5$  come stima per i passi MC necessari all'equilibratura del sistema, ma andando ad analizzare più nel dettaglio la situazione scopriamo che questo non è sufficiente quando il sistema inizia a nematizzare. Quando infatti la pressione è abbastanza alta per consentire la transizione di fase, lentamente il sistema inizia a diminuire ulteriormente il suo volume, aumentando di conseguenza la "volume fraction"  $\varphi$  del sistema.

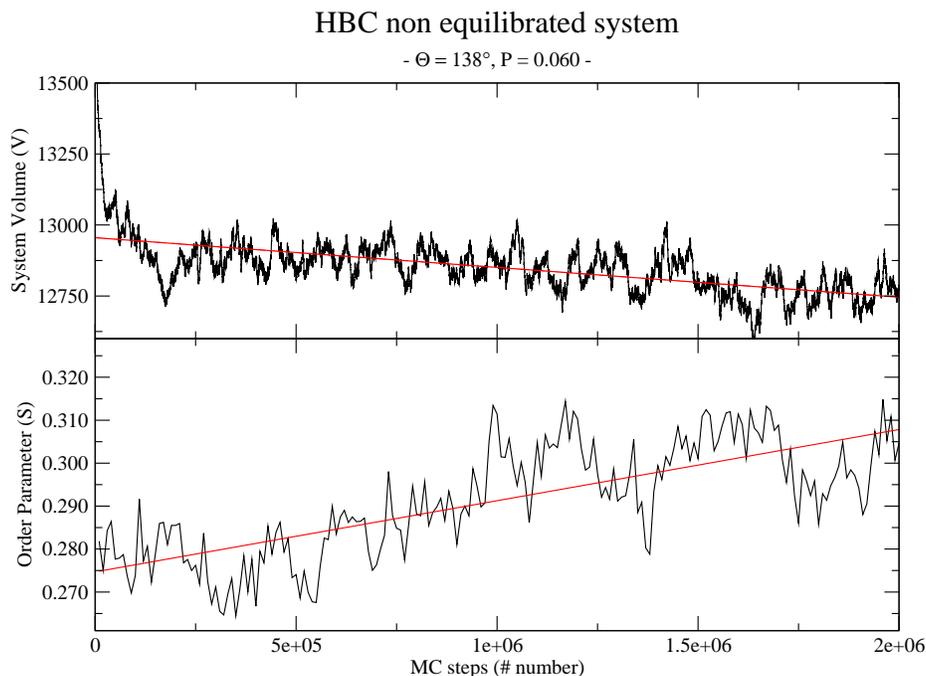


Figura 13.2: Controllo parallelo del volume e del parametro d'ordine di un sistema ad alta piegatura,  $\Theta = 138^\circ$ , in cui risulta evidente che i  $2 \cdot 10^6$  step MC (dopo i precedenti  $4 \cdot 10^6$ ) non sono ancora sufficienti a farci raggiungere una configurazione stabile. Questo fatto ci impedirà di visualizzare graficamente la regione di transizione quando andremo a graficare i diagrammi di fase del secondo gruppo di HBC.

Tale effetto è ben visibile nel secondo run che mostra come, da una situazione di apparente stabilità, il sistema stia ancora evolvendo verso stati di equilibrio nematico. Nel grafico a destra della figura 13.1 abbiamo riportato a titolo di esempio la situazione degli HBC con angolarità  $\Theta = 150^\circ$  che risultano nematizzare alla pressione di  $P = 0.019$ , nel quale è evidente come anche dopo  $4 \cdot 10^6$  passi MC il sistema non abbia raggiunto una condizione di volume di equilibrio. In questo caso il fit lineare ci aiuta a visualizzare come il volume del sistema stia ancora diminuendo e ciò si ripercuote concretamente nell'impossibilità di osservare la regione di coesistenza nel diagramma di fase.

E' ancora peggiore la situazione per quanto riguarda angolarità minori, quando i doppi cilindri sono più piegati è più difficile riuscire a stimare la  $\varphi_N$  poiché l'equilibratura del sistema nematico richiede molti più step MC per essere raggiunta.

In particolare, per quanto riguarda il sistema più piegato con  $\Theta = 138^\circ$  per cui riusciamo ad osservare una transizione di fase nel range di pressioni osservato, in figura 13.2 possiamo vedere come dopo  $4 \cdot 10^6$  passi MC in compressione e dopo ulteriori  $2 \cdot 10^6$  il sistema non abbia ancora stabilizzato il suo volume. E' però evidente che in questa circostanza, a pressione

$P = 0.060$ , il sistema stia nematizzando, anche perché parallelamente aumenta il parametro d'ordine come indicato approssimativamente dai fit lineari riportati su ciascun grafico.

Nel seguito terremo sempre conto di ciò, esaminando l'andamento dei volumi e dei parametri d'ordine per ciascun punto riportato in figura, e da questa analisi potremo stabilire come varia la regione di coesistenza al variare del “bending” in termini della “volume fraction”. Insisteremo in particolare come i valori ottenuti per angoli inferiori a  $\Theta < 150^\circ$  siano delle sottostime della  $\varphi_N$  dato che l'equilibratura non è stata ancora raggiunta, ma nonostante ciò saremo in grado di ottenere dei risultati concreti anche confortati dall'andamento della  $\varphi_I$  dove inizia la transizione di fase. Proprio per come è stata strutturata l'analisi è molto più semplice stimare dove la transizione inizia piuttosto che dove è completata, per questo anche la “volume fraction” della fase isotropica sarà un parametro interessante da controllare.

## 13.2 Il primo gruppo

Cominciamo la nostra analisi con lo studio del comportamento di un sistema formato da doppi cilindri poco piegati, prendendo in esame per il primo gruppo gli HBC costruiti con apertura maggiore:

$$\Theta_{HBC}^{150^\circ} = 150^\circ \quad \Theta_{HBC}^{155^\circ} = 155^\circ \quad \Theta_{HBC}^{160^\circ} = 160^\circ. \quad (13.1)$$

Questi valori sono stati scelti in linea con quanto ottenuto dall'analisi del “bending” fosforico, con l'intento di esplorare un range abbastanza vasto di valori di piegatura per poter includere anche gli errori riscontrati dal calcolo delle osservabili dinamiche definite in precedenza.

Decidiamo di cominciare la nostra analisi con valori abbastanza vicini al caso rettilineo per poter poter estendere concettualmente quanto visto nel caso di singoli cilindri in assenza di “bending”. Inoltre questo fatto ci permetterà visivamente una migliore osservazione della regione di coesistenza: essendo costituito da oggetti poco piegati il sistema non avrà difficoltà particolari nel transire nella fase nematica necessitando di pochi step MC per raggiungere la sua fase stabile come già discusso in fase di equilibratura.

In figura 13.3 sono riportati i diagrammi di fase ottenuti per i tre gradi di piegatura delle sequenze considerate. Come possiamo vedere la regione di coesistenza della sequenza meno piegata delle tre è molto ben definita sia in fase di compressione che di decompressione del sistema, mentre per la sequenza intermedia cominciamo a vedere qualche effetto dovuto alla stabilizzazione più lenta. Nel grafico della sequenza più piegata la discontinuità del diagramma di fase comincia ad essere molto meno visibile, fatto sicuramente dovuto al gran numero di passi Monte Carlo necessari a far transire il sistema completamente.

Da una prima analisi visiva salta subito agli occhi che la “volume fraction” critica di nematizzazione aumenta all'aumentare del grado di piegatura mentre il valore del parametro

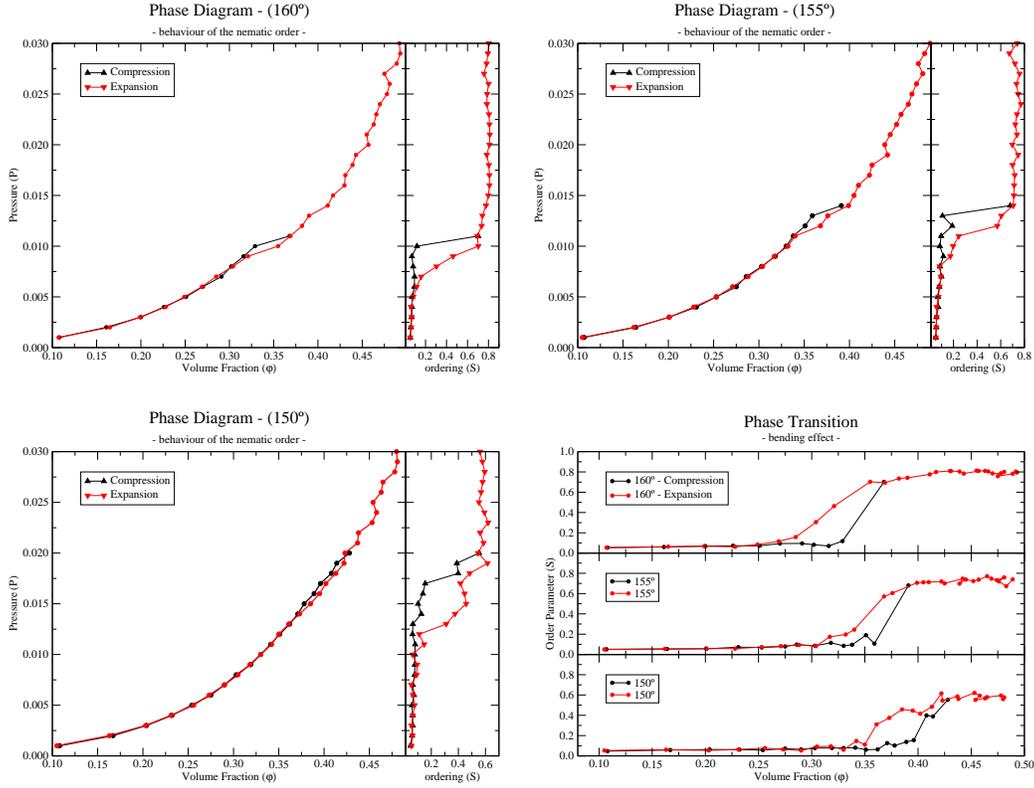


Figura 13.3: Diagrammi di fase ottenuti dalle simulazioni MC in compressione e decompressione di sistemi appartenenti al primo gruppo di piegature. Notiamo come sia ben evidente la regione di coesistenza delle fasi isotropiche e nematiche per i due HBC meno piegati, mentre nel terzo caso il sistema sta transendo pur non avendo completamente raggiunto la situazione stabile.

d'ordine della fase nematica diminuisce sensibilmente. Questo è in linea con quanto previsto, infatti una struttura piegata formerà degli aggregati rettilinei con difficoltà sempre maggiore all'aumentare del grado di piegatura. Inoltre in linea di principio sono possibili delle configurazioni circolari chiuse che via via diventano sempre più probabili all'aumentare del "bending" poiché diminuisce il numero di HBC necessari alla chiusura di tale struttura. Questo fenomeno è sicuramente dannoso alla nematizzazione e abbassa in modo via via più rilevante il parametro d'ordine della transizione stessa disallineando i vettori introdotti nella 12.12 con cui abbiamo definito il calcolo di tale grandezza.

Dall'analisi dei risultati ottenuti possiamo stimare le concentrazioni medie a cui il sistema transisce:

$$\varphi_I^{160^\circ} = 0.302 \pm 0.008 \quad \varphi_N^{160^\circ} = 0.355 \pm 0.013 \quad (13.2)$$

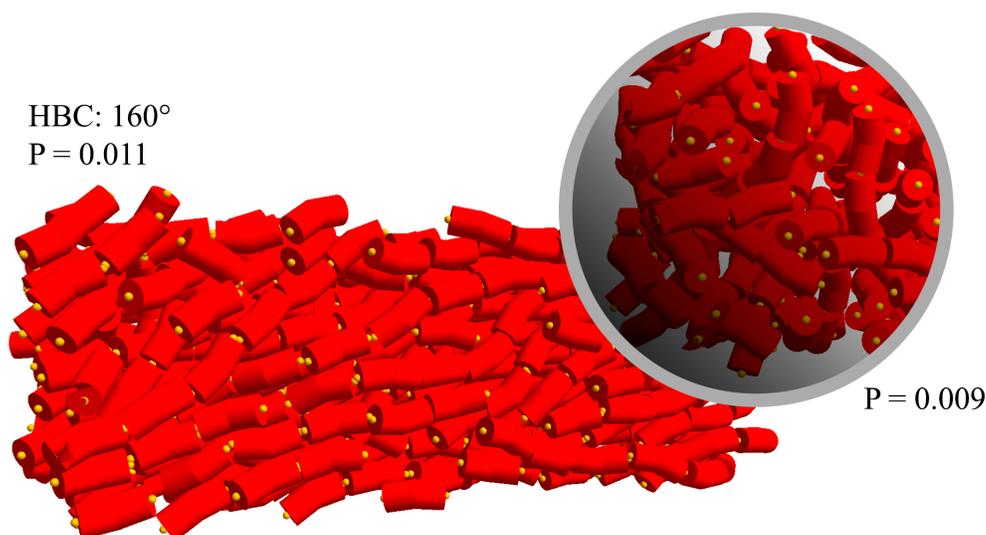


Figura 13.4: Illustrazione della situazione riscontrata nel caso  $\Theta^{160^\circ}$  a pressioni diverse. L'asse di nematizzazione è evidentemente lungo una diagonale del parallelepipedo che contiene gli oggetti.

$$\varphi_I^{155^\circ} = 0.338 \pm 0.019 \quad \varphi_N^{155^\circ} = 0.376 \pm 0.008 \quad (13.3)$$

$$\varphi_I^{150^\circ} = 0.371 \pm 0.016 \quad \varphi_N^{150^\circ} = 0.422 \pm 0.011 \quad (13.4)$$

dove gli errori sono stati presi facendo la semidifferenza dei valori massimi e minimi a cui il sistema effettuava la transizione di fase durante la fase di compressione ed espansione.

Attraverso l'introduzione del "bending" perciò siamo riusciti ad aumentare la concentrazione critica a cui il sistema nematizza. Il caso rettilineo con  $\beta\Delta E_s = 8.33$  fornisce  $c_N^{180^\circ} = 316 \text{ mg/ml}$  per i valori teorici della concentrazione nematica mentre nel caso a maggiore piegatura arriviamo ad una concentrazione equivalente di  $c_N^{150^\circ} = 406 \text{ mg/ml}$ . Con una piegatura di  $30^\circ$  rispetto al caso di cilindro rettilineo e riscaldando il volume degli HBC abbiamo introdotto una variazione pari a circa il 29% della concentrazione critica di nematizzazione. Tutto questo ancora non è sufficiente a spiegare il grande divario tra le concentrazioni ottenute al variare delle sequenze nell'articolo di Nakata [16], ma è ragionevole supporre che tale fenomeno abbia un comportamento estremamente non lineare. La non linearità del fenomeno è dovuta in gran parte alla dipendenza esponenziale di  $M$ , lunghezza media degli aggregati, al variare della "volume fraction"  $\varphi$  come osservato da De Michele [19], poiché qualitativamente possiamo dire che più le catene sono lunghe e più probabilità ci sono che esse non siano lineari. Per cercare di comprendere meglio la situazione introdotta dal "bending" decidiamo di analizzare il secondo

gruppo di HBC, andando ad esplorare il comportamento a gradi di piegatura più elevati.

### 13.3 Il secondo gruppo

Il secondo gruppo che andremo a studiare richiede un'analisi più accurata dei dati ed un setup sperimentale più fine a causa dei gradi di piegatura maggiori rispetto al caso precedente:

$$\Theta_{HBC}^{145^\circ} = 145^\circ \quad \Theta_{HBC}^{141^\circ} = 141^\circ \quad \Theta_{HBC}^{138^\circ} = 138^\circ \quad \Theta_{HBC}^{134^\circ} = 134^\circ. \quad (13.5)$$

A queste angolature, come già discusso, il sistema necessita di un grandissimo numero di mosse Monte Carlo prima di raggiungere la densità di stabilizzazione del sistema. Alla luce di questo fatto decidiamo di procedere in modo analogo al precedente, effettuando due gruppi di run da  $2 \cdot 10^6$  step MC consecutivi comprimendo il sistema e successivamente decomprimendolo per altri  $2 \cdot 10^6$  passi, partendo da quella che abbiamo supposto essere la prima configurazione nematica raggiunta: anche se non avremo raggiunto una completa nematizzazione saremo in grado di valutare come l'aumento del “bending” influisca sulla transizione di fase del sistema.

In realtà in questo caso la stima è piuttosto approssimativa, dato che nel diagramma di fase di tutti i sistemi considerati non riusciamo ad individuare graficamente la regione di coesistenza delle fasi. Per poter stimare la “volume fraction” di transizione occorrerà basarci principalmente sull'analisi dell'andamento del parametro d'ordine supponendo il sistema nematizzato nel primo punto in cui, dopo la discontinuità di  $S$ , esso si attesta ad un valore circa costante.

In questo caso in figura 13.6 decidiamo di riportare solamente i diagrammi di fase delle simulazioni in decompressione poiché ottenuti in totale dopo  $6 \cdot 10^6$  passi MC e quindi maggiormente equilibrati rispetto a quelli in compressione, nonostante anche in questo caso non venga osservata la regione di transizione.

Da quanto ottenuto, risulta evidente che il sistema con  $\Theta_{HBC}^{145^\circ} = 145^\circ$  sta nematizzando in maniera abbastanza evidente ed anche

HBC: 134°  
P = 0.030

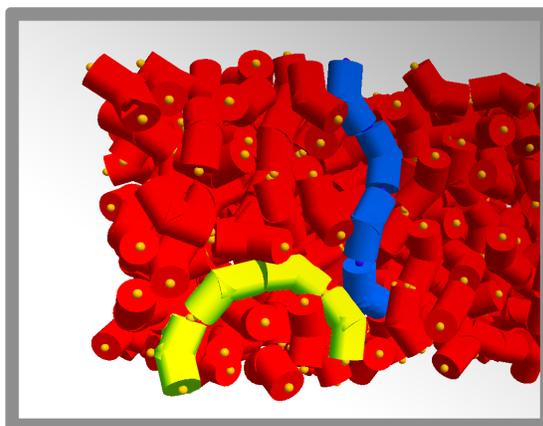


Figura 13.5: Particolare della formazione di due tipi di aggregati nel caso di sistemi a “bending” maggiore. In verde è rappresentata la situazione peggiore prevista, quella in cui l'aggregazione chiude su se stessa distruggendo il bending, mentre in blu è rappresentata la formazione di un aggregato quasi rettilineo.

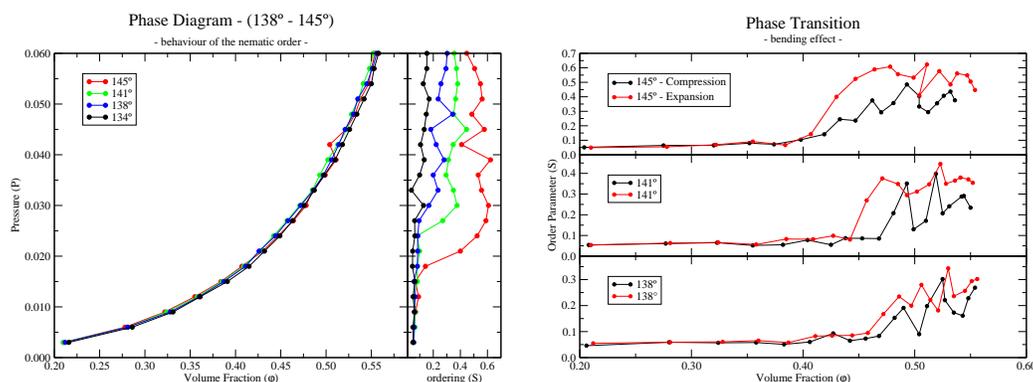


Figura 13.6: Diagramma di fase del secondo gruppo di piegature degli oggetti HBC. In questo caso la discontinuità indotta dalla transizione di fase non è mai evidente e la zona in cui il sistema transisce è possibile stimarla solamente attraverso l’analisi della discontinuità del parametro d’ordine.

la regione di coesistenza sembra caratterizzata da una grande ampiezza della zona metastabile. Una situazione più o meno analoga si ripropone anche nel caso di “bending” inferiore,  $\Theta_{HBC}^{141^\circ} = 141^\circ$ , mentre ad aperture inferiori osserviamo che il parametro d’ordine comincia la sua discontinuità intorno a valori piuttosto alti della “volume fraction” ( $\varphi > 0.5$ ), raggiungendo valori inferiori ad  $S = 0.5$  che avevamo considerato essere la soglia minima affinché il sistema fosse transito in una fase nematica. Questo comporta che la pressione a cui il sistema nematizza deve essere superiore al massimo valore  $P = 0.060$  considerato in questa sede e le corrispondenti stime “volume fraction” ottenute possono essere solamente indicative del valore inferiore delle densità reali a cui il sistema transisce. Per convincersi di questo fatto risulta particolarmente utile l’analisi dell’andamento del volume del sistema, e parallelamente della  $S$ , fatta in precedenza nella sezione di equilibratura e riportata in figura 13.2.

Alla luce di ciò possiamo stimare i valori critici della frazione di volume che definisce la regione di coesistenza per tre dei sistemi del secondo gruppo, ottenendo i valori:

$$\varphi_I^{145^\circ} = 0.419 \pm 0.012 \quad \varphi_N^{145^\circ} = 0.478 \pm 0.014 \quad (13.6)$$

$$\varphi_I^{141^\circ} = 0.460 \pm 0.014 \quad \varphi_N^{141^\circ} = 0.528 \pm 0.018 \quad (13.7)$$

$$\varphi_I^{138^\circ} = 0.504 \pm 0.014 \quad \varphi_N^{138^\circ} = 0.556 \pm 0.019 \quad (13.8)$$

mentre nulla riusciamo a prevedere per la “volume fraction” nematica degli HBC più piegati. In realtà possiamo pensare di stimare per la fase isotropica il valore approssimato di  $\varphi_I^{134^\circ} = 0.548$ , valore nel quale il parametro d’ordine sembra iniziare la discontinuità caratteristica della transizione.

Anche in questo caso possiamo calcolare le corrispondenti concentrazioni critiche ottenendo che la transizione di fase a più alta densità avviene a  $c_N^{138^\circ} = 511 \text{ mg/ml}$ , introducendo quindi una variazione del 61% nella concentrazione rispetto al caso teorico di cilindri non piegati.

### 13.4 Dipendenza della concentrazione critica dalla conformazione della molecola

Dai dati ricavati in precedenza attraverso lo studio di ciascuno dei sistemi considerati possiamo pensare di analizzare il ruolo giocato dalla conformazione del modello “coarse grained” introdotto per descrivere il doppio filamento corto di DNA.

Per definire le regioni di coesistenza delle fasi grafichiamo la volume fraction ottenuta rispetto al grado di “bending” introdotto, riportando per completezza i valori teorici [19] previsti nel caso di cilindro rettilineo (HC) con “aspect ratio” pari a  $X_0 = 2.15$ .

Dai punti rossi riportati sul diagramma in figura 13.8 si evince chiaramente il comportamento non lineare delle “volume fraction” critiche riscontrato al variare del “bending” dei doppi cilindri HBC trattati in questa sede. Come previsto l’introduzione della piegatura nella conformazione strutturale degli oggetti cilindrici ha indotto il sistema a transire a densità via via sempre più elevate con l’aumentare del grado di piegatura.

Da quanto emerso dall’analisi dati e dalla verifica di ciascuna delle configurazioni ottenute siamo riusciti a risalire al comportamento del sistema generato dalla rottura della simmetria di rotazione azimutale propria della struttura cilindrica. I valori di ciascuna delle “volume fraction” ottenute riportate nella figura derivano dall’analisi statistica di un gran numero di simulazioni MC di sistemi “bent-cylinders”.

Per quanto riguarda la curva che delimita la regione isotropica gli errori riscontrati sono inferiori alle dimensioni del simbolo usato nel diagramma, in questo caso infatti la regione risulta ben definita per ogni sistema a diversa piegatura e può essere stimata non appena il sistema comincia a subire incrementi significativi nel valore del parametro d’ordine. Alla luce di quanto osservato in tutti i run eseguiti, la fase isotropica è in generale sempre caratterizzata da valori  $S \simeq 0.06$  ed un aumento repentino del 100% di tale valore fornisce l’indicazione del punto in cui il sistema comincia a formare aggregati lineari con lunghezza rilevante. Abbiamo supposto come limite della regione isotropica i valori della “volume fraction” per cui il parametro d’ordine si mantiene sempre inferiore a  $S < 0.2$  e successivamente mediato tra le simulazioni in compressione e decompressione per depurare il sistema da eventuali fenomeni di metastabilità<sup>1</sup>. Questo ci ha permesso di delimitare con precisione la regione isotropica

<sup>1</sup>Durante la fase di compressione il sistema fatica a transire nella fase nematica portandosi in condizione di sovra-saturazione, viceversa nella fase di decompressione in cui il sistema si trova in condizioni di sotto-saturazione.

ed al tempo stesso fornisce ovviamente un limite inferiore per la regione nematica, che dovrà necessariamente trovarsi al di sopra di essa oltre la regione di coesistenza.

A differenza della regione isotropica, le “volume fraction” che caratterizzano la curva di fase nematica sono decisamente delle stime inferiori dei valori effettivi che raggiungerà il sistema una volta completata la transizione di fase. Solo per gli HBC meno piegati,  $\Theta^{160^\circ}$  e  $\Theta^{155^\circ}$ , riusciamo ad individuare graficamente la discontinuità del diagramma di fase del sistema, chiaro indice di una completa nematizzazione, supportata dalla discontinuità del parametro d’ordine che raggiunge valori superiori a  $S > 0.7$  in entrambi i casi. Il fatto saliente è che, avendo periodicamente controllato le simulazioni, la discontinuità della curva di diagramma di fase comincia ad essere netta e ben visibile quando il parametro d’ordine raggiunge valori superiori a  $S > 0.6$ , infatti nella prima serie di  $2 \cdot 10^6$  run la discontinuità dell’isoterma non era evidente nemmeno nelle sequenze più rettilinee e solo raddoppiando la durata di questi run siamo riusciti ad apprezzarla. A causa di questo fatto con gradi di piegatura superiori la transizione resta visibile esclusivamente grazie alla discontinuità del parametro d’ordine, che tra l’altro si fa sempre meno marcata con l’aumentare del “bending” delle strutture a doppio cilindro. Per questi motivi i valori ottenuti per la “volume fraction”, specialmente per gradi al di sotto di  $\Theta < 145^\circ$ , sono sicuramente delle sottostime delle densità a cui il sistema nematizza, infatti il mancato raggiungimento della soglia critica per il parametro d’ordine ( $S > 0.6$ ) non ci consente di apprezzare il salto di densità nel diagramma di fase che caratterizza la termodinamica di tali sistemi.

Basandoci su quanto emerso dal comportamento del parametro d’ordine abbiamo riportato delle stime della  $\varphi_N$  osservando la pressione a cui il sistema in compressione iniziava a mostrare una discontinuità di  $S$  e quando in espansione la terminava riportandosi nello stato isotropico. In realtà per ottenere delle stime realistiche in decompressione siamo partiti dalla prima configurazione nematica ottenuta diminuendo la pressione: in questo modo abbiamo osservato la pressione a cui il sistema trovava conveniente distruggere gli aggregati lineari diminuendo in modo sensibile il valore del parametro d’ordine. Partendo da un determinato stato nematico, infatti, il primo punto per il quale  $S$  non diminuisce in modo significativo sarà il punto del diagramma delle fasi in cui la nematizzazione è più probabile e la corrispondente densità del sistema sarà la nostra  $\varphi_N$ . Poiché abbiamo ottenuto la “volume fraction” in decompressione, a causa della metastabilità, essa sarà una sottostima del valore effettivo di tale grandezza, errore che si aggiunge alla mancata presenza della discontinuità dell’isoterma (che innalzerebbe ulteriormente la densità di nematizzazione). Ad aiutarci ulteriormente a comprendere dove il sistema stesse transendo abbiamo sfruttato l’interpolazione di Bezier<sup>2</sup>, che individua il comportamento del parametro d’ordine consentendoci di valutare la regione in cui si manifesta la discontinuità. In figura 13.7 è possibile visualizzare l’andamento interpolato di  $S$  per ciascun

---

<sup>2</sup>Le spline di Bezier sono sfruttate dai moderni software di grafica (caratteri TrueType, Adobe Flash, GIMP, PostScript) per l’interpolazione di punti a geometria curvilinea.

sistema HBC da cui è visibile come la transizione di fase si sposti via via a pressioni, e quindi a “volume fraction”, superiori aumentando la piegatura del sistema.

Alla luce di tutto ciò la situazione risulta comunque evidente, come aspettato, che la densità a cui il sistema trova conveniente formare aggregati lineari comincia ad innalzarsi, vincolata anche dall’andamento del limite della regione isotropica, all’aumentare del grado di piegatura dei singoli “building blocks” del sistema, raggiungendo una variazione massima di almeno il 57% nell’intero range di  $22^\circ$  esplorato in questa tesi. La variazione della “volume fraction” isotropica invece risulta essere addirittura del 67% rispetto al caso meno piegato, un valore che ben si rapporta alla variazione del 70% nella concentrazione critica ottenuta riscontrata negli esperimenti di Zanchetta e Nakata [17, 16].

L’obiettivo che ci eravamo proposti in questa sede non era esattamente di fornire un modello quantitativo in grado di prevedere comportamenti dei sistemi termodinamici considerati, quanto di mostrare come l’introduzione di una rottura di simmetria azimutale di un sistema cilindrico attraverso la piegatura lungo l’asse si ripercuotesse sulla sua capacità di aggregazione, andando ad influenzare criticamente la termodinamica del sistema modificando la “volume fraction” critica di nematizzazione. Possiamo pensare di congiungere i risultati ottenuti nella prima e nella seconda parte per vedere le correlazioni tra le concentrazioni critiche di nematizzazione ed i gradi di piegatura studiati, per valutare dove il nostro modello funziona e come eventualmente migliorarlo dal punto di vista sia della definizione del “bending” molecolare sia nella modellizzazione stessa degli oggetti HBC.

A causa della loro definizione, i doppi cilindri sono caratterizzati dall’aver un volume  $V_{HBC}$  via via maggiore all’aumentare della piegatura, mentre avevamo supposto che il volume effettivo della molecola riscontrato nelle simulazioni MD dovesse diminuire man mano che la sequenza piegava. Nel lavoro di De Michele [19] avevamo visto che l’effetto legato all’aumento del valore dell’“aspect ratio”  $X_0$  nel caso di cilindri o superquadriche, riportato in figura 13.11, e di conseguenza del volume del singolo “building block” porta ad una diminuzione non trascurabile delle “volume fraction” a cui il sistema transisce. Questo fatto ancora una volta ci suggerisce che stiamo sottostimando le concentrazioni critiche in modo via via maggiore con l’aumentare del grado di piegatura e possiamo pensare di tenerne conto anche nel nostro caso facendo alcune considerazioni.

Pensiamo di valutare la differenza nell’andamento delle concentrazioni critiche dapprima

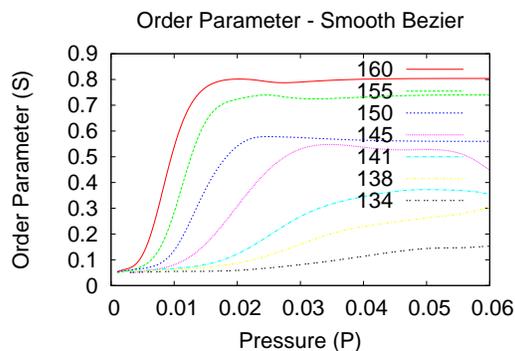


Figura 13.7: Interpolazione Bezier ottenuta con gnuplot per l’andamento del parametro d’ordine in funzione della pressione.

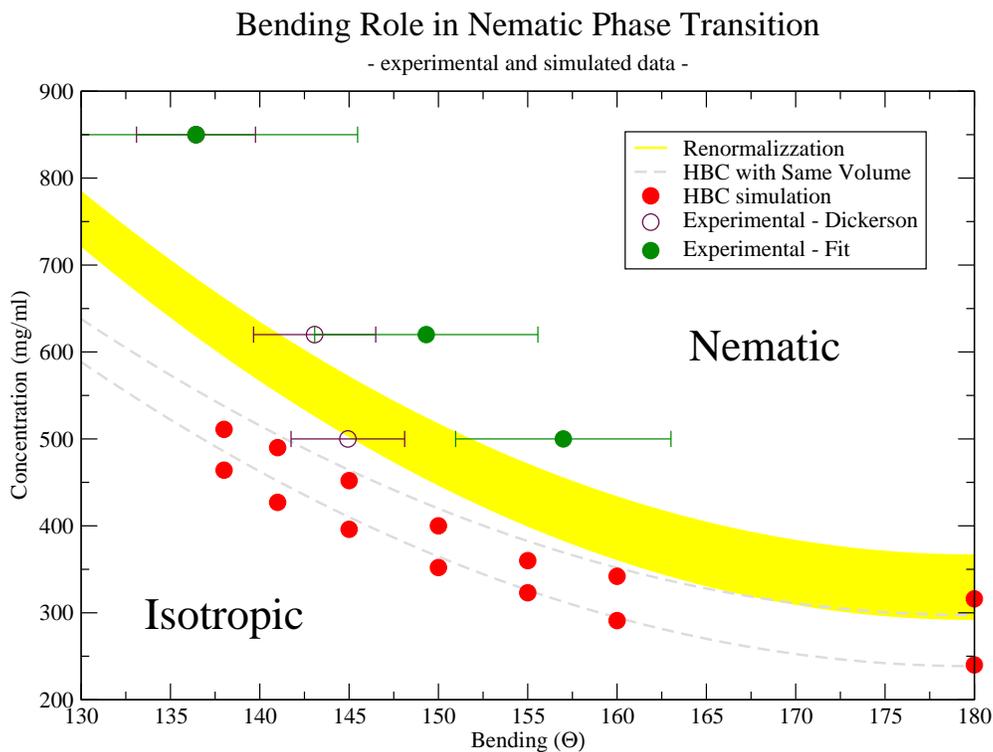


Figura 13.8: Ipotesi sulla regione di coesistenza se gli HBC avessero tutti lo stesso volume del cilindro rigido con  $X_0 = 1.75$ . Il diagramma è puramente indicativo, mentre i punti rossi rappresentano la regione di coesistenza riscontrata nelle nostre simulazioni MC.

usando i volumi calcolati per i singoli HBC e successivamente supponendo che tutti i doppi cilindri abbiano sempre lo stesso volume del caso rettilineo con  $X_0 = 2.15$ . In questo modo otteniamo la curva tratteggiata in figura 13.8 che mostra come varia la curva di coesistenza isotropica-nematica ipotizzando di introdurre un modello HBC a volume costante. Inoltre ricordandoci che in realtà le sequenze di 12 basi azotate possiedono inizialmente un “aspect ratio” inferiore a quanto visto, pari a circa  $X_0 = 1.75$ , possiamo riscalarne linearmente la curva appena ottenuta, per valutare la situazione determinata da una ulteriore diminuzione del volume del singolo cilindro (ricordando che la diminuzione è introdotta solo nella lunghezza del cilindro, non nel diametro delle basi). Tale ipotetica situazione è rappresentata dalla regione gialla nella figura 13.8.

Grazie a semplici considerazioni abbiamo raggiunto le grandezze tipiche delle concentrazioni critiche riportate negli esperimenti di Nakata [16], anche se purtroppo i valori per il “bending” delle osservabili analizzate mediante le simulazioni a MD non ci consentono di mappare correttamente i valori sperimentali delle tre sequenze (AAT, AAC ed ACC) con la regione di coesistenza ottenuta. Possiamo solamente notare che la sequenza ACC che tran-

sisce a concentrazioni di  $850\text{mg/ml}$  sembra essere meglio rappresentata dal valore ottenuto con l'osservabile di Dickerson [44], mentre la più rettilinea AAT segue un accordo migliore con quello legato al "bending" fosforico. Questa piccola digressione è una pura teorizzazione degli effetti ottenibili studiando HBC più corti e di pari volume e non tiene assolutamente conto della presenza di eventuali altri fenomeni, ma mostra quanto vaste siano ancora le possibilità di studio connesse al fenomeno del piegamento delle strutture delle molecole di DNA.

In definitiva abbiamo analizzato 30 valori di pressione per ciascun grado di piegatura degli HBC del primo gruppo e 20 per quelli del secondo, sfruttando  $4 \cdot 10^6$  passi Monte Carlo per le simulazioni in compressione e  $2 \cdot 10^6$  in decompressione. Per effettuare ciascuno di questi run sono stati necessari almeno 5 giorni di esecuzione del programma di simulazione MC `ellipsHC` sviluppato dal prof. De Michele analogo a quello usato per ricavare i diagrammi di fase dei sistemi a cilindri rigidi o a superquadriche, modificato per adattarlo alla nuova geometria a doppio cilindro piegato.

Nonostante la grande mole di dati analizzati e gli accorgimenti utilizzati non siamo riusciti ad osservare la piena nematizzazione dei sistemi maggiormente piegati, fatto che attribuiamo in gran parte al contributo entropico sfavorevole introdotto dalla rottura della simmetria azimutale del cilindro rettilineo. Anche graficamente, ma in particolare studiando la tipologia di aggregati presenti in alcuni frame delle simulazioni maggiormente piegate, abbiamo riscontrato più volte la formazione di aggregati a forma di anello, supposti in via teorica in figura 12.3 e riscontrati nella maggior parte delle simulazioni ad alta pressione. Nella figura 13.5 riportiamo una configurazione presa da una simulazione di un sistema formato da HBC a "bending" maggiore,  $\Theta^{134}$ , in cui è mostrata chiaramente la coesistenza di aggregati lineari con altri di forma anulare. Questi aggregati, piuttosto stabili dal punto di vista meccanico, determinano un contributo che ostacola la transizione del parametro d'ordine creando stati metastabili particolarmente fastidiosi. In seguito ci proponiamo perciò<sup>3</sup> di analizzare sistemi HBC usando un numero molto più grande di passi MC, almeno pari a  $10^7$ , per favorire il processo di nematizzazione ed ottenere delle stime più precise delle densità a cui il sistema completa la transizione di fase. Facendo ciò potremo riscalarare la curva che rappresenta la regione di coesistenza e calcolarne computazionalmente l'andamento corretto.

---

<sup>3</sup>Nel momento in cui stiamo scrivendo abbiamo già lanciato ulteriori simulazioni per cercare di svelare la transizione di fase degli HBC ad alta piegatura.



# Conclusioni

Il controllo di una gran quantità di interazioni tipiche della “soft matter” è uno degli obiettivi primari dello studio della Materia Condensata che, oltre a fornire una spiegazione fisica dei fenomeni ad esse connesse, ci mette a disposizione uno strumento estremamente versatile per la realizzazione di materiali innovativi.

In particolare tra le transizioni termodinamiche più interessanti prese in esame dalla fisica della materia soffice, troviamo la formazione di fasi liquido-cristalline, che consentono attraverso il controllo di alcuni semplici parametri settati a priori, la riorganizzazione controllata del sistema fino ad ordini di grandezze tipiche delle dimensioni molecolari. Molte ragioni si nascondono dietro tale transizione di fase, dal riordinamento di molecole causato da effetti legati al volume escluso [14, 15] fino al considerare interazioni tra gli stessi costituenti attraverso forze di “stacking” a carattere idrofobico [17, 16, 12, 19, 21] che determinano i processi di “self assembly” nel sistema.

Naturalmente tale fenomeno assume un interesse maggiore quando come “building block” per la formazione di fasi nematiche viene preso in esame il doppio filamento di DNA: grazie a processi di auto organizzazione dell’acido nucleico è stato possibile formare origami come fatto da Rothmund [9] ed ha permesso al team di Luo realizzare scritte in hydrogel [7, 8], ma anche controllare transizioni di fase funzionalizzando il doppio filamento su particelle colloidali come ideato da Mirkin [11] e successivamente studiato nel dettaglio da Bellini [10] e da Glotzer [12]. Tutto questo offre sempre nuove prospettive di sviluppo nello studio dei fenomeni di aggregazione e nella realizzazione di nuovi materiali dalle caratteristiche innovative, perciò la comprensione dei fenomeni che si celano dietro questi particolari comportamenti è sempre più interessante ed oggetto di studi sempre più approfonditi.

Parallelamente alla ricerca del controllo sperimentale dei fenomeni legati a queste transizioni di fase, la crescente potenza di calcolo messa a disposizione dei ricercatori ha permesso la simulazione del comportamento dinamico e termodinamico di tali sistemi, offrendo un modo estremamente efficace per osservare come il comportamento del singolo costituente influisca nel fenomeno di “self assembling” permettendo l’ordinamento su grande scala dell’intero sistema.

Per quanto riguarda la transizione di fase nematica del doppio filamento di DNA, la spiegazione del fenomeno è stata attribuita in gran parte agli effetti di volume escluso delle

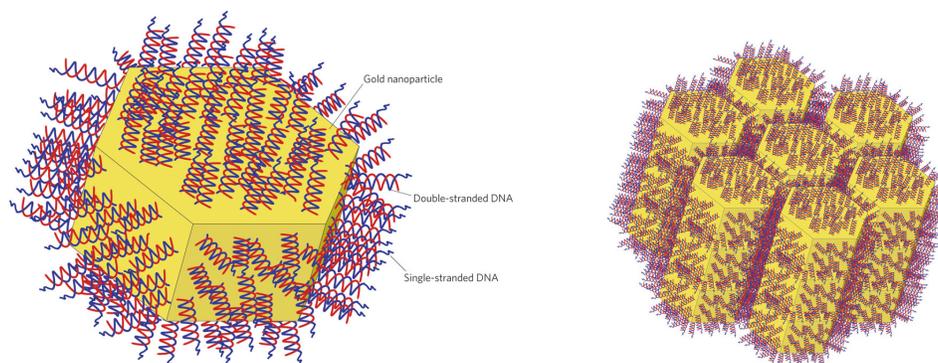


Figura 13.9: Funzionalizzazione di DNA su nanoparticelle di oro e relativa formazione dell'aggregato. Immagini prese dal lavoro di Glotzer [12].

particelle, ovvero la regione di volume inaccessibile a tutte le altre molecole del sistema, se la lunghezza di persistenza dell'acido nucleico fosse abbastanza grande da poterlo considerare rettilineo. Attraverso il modello OBF [14, 15] è stato possibile mettere in evidenza fenomeni di formazioni di fasi liquido cristalline per tutte quelle molecole la cui forma può essere approssimata ad un cilindro rettilineo allungato, permettendo di prevedere le densità a cui il sistema nematizza e caratterizzando completamente la termodinamica del sistema.

La previsione teorica dell'esistenza di un limite inferiore nell' "aspect ratio"  $X_0$  del singolo costituente, determinata dallo studio di questo modello e riscontrata sperimentalmente, poneva una soglia al di sotto della quale tali molecole potessero auto organizzarsi grazie a soli effetti di volume escluso, impedendo di fatto la formazioni di fasi liquido cristalline. Tale affermazione venne in seguito smentita dai lavori di Nakata e Zanchetta [17, 16] nel 2007, i quali disidratando sistemi composti da doppi filamenti di DNA corti ritrovano la transizione di fase al di sotto di valori critici di  $X_0 \simeq 3$ , mostrando inoltre ulteriori caratteristiche salienti come ad esempio la formazione di una fase nematica a chiralità ben definita.

La spiegazione di questo nuovo fenomeno, come già dettagliatamente discusso, è stata attribuita alla presenza di interazioni di "stacking" idrofobico tra le basi terminali dell'acido nucleico. Questo fenomeno si è rivelato di grande interesse scientifico nella fisica della materia soffice, poiché introduce un nuovo strumento per controllare i fenomeni di "self-assembly" di sistemi di DNA attraverso le interazioni legate allo "stacking end-to-end" tra le basi estremali dei singoli "building block" del sistema. Tale fenomeno si presta ad interessanti applicazioni tecnologiche ed ulteriori caratteristiche dell'interazione tra sistemi colloidali mediata da filamenti di DNA è stata studiata in grande dettaglio da Bellini et al. [10].

Grazie all'ausilio di numerose tecniche di simulazione dinamica è stato possibile stimare le grandezze tipiche di tali interazioni sia sfruttando la dinamica molecolare, come fatto nel lavoro di Maffeo [21] in cui un modello "full atom" che interagisce attraverso un "force field" AMBER ha consentito di stimare le forze di adesione tra basi terminali di due segmenti corti

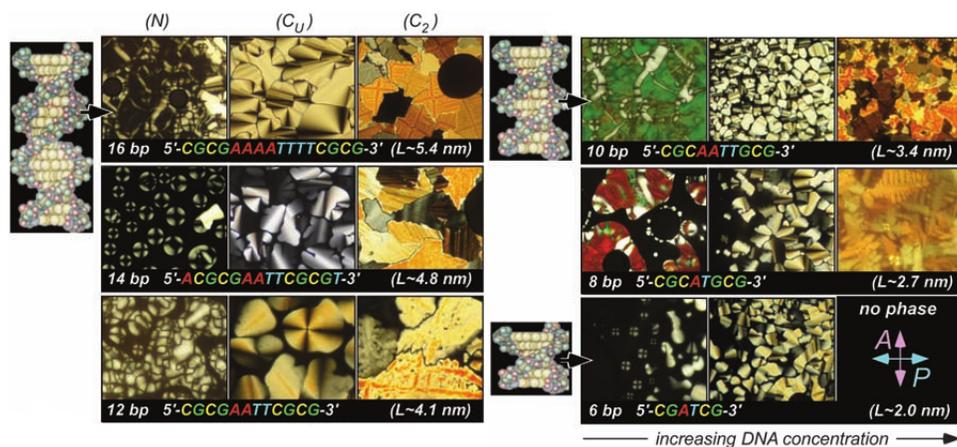


Figura 13.10: Evidenza sperimentale della transizione di fase LC per segmenti di DNA con  $X_0 < 3.0$  riscontrata nel lavoro di Nakata et al.[16].

di DNA, sia servendoci di un approccio statistico offerto dalle simulazioni Monte Carlo. Tale tipo di simulazione necessita la creazione di un modello “coarse grained” che rappresenta il DNA per poter sondare lo spazio delle fasi del sistema: nel lavoro di Glaser [25] vengono trattati sistemi contenenti cilindri corti (HC) o dischetti dotati di siti attrattivi su ciascuna base, mentre De Michele [19] ha proposto uno studio attraverso l’utilizzo di superquadriche di forma cilindrica (SQ).

In questo contesto si è svolto il nostro lavoro di tesi, con l’intento di fornire una nuova interpretazione dei possibili fattori implicati nella transizione di fase indotta durante il processo di “self assembly”. I modelli termodinamici con siti attrattivi posti agli estremi dell’oggetto proposti in precedenza, siano essi schematizzabili mediante cilindri rigidi (HC) o superquadriche di forma cilindrica (SQ) [19, 25], hanno fornito dei buoni risultati per quanto riguarda la stima delle forze di “stacking”, mentre falliscono nel prevedere la concentrazione critica di nematizzazione al variare della sequenza. In entrambi i modelli la “volume fraction” di nematizzazione risulta indipendente dal tipo di sequenza considerata e caratterizzata esclusivamente dal rapporto di forma geometrico  $X_0 = \frac{L}{D}$  (“aspect ratio”). Gli esperimenti di Nakata e Zanchetta [17, 16] infatti, non solo portano alla luce il fenomeno di formazione di una fase liquido cristallina in soluzioni concentrate di DNA, ma mostrano anche una forte dipendenza della concentrazione di nematizzazione dalla sequenza dell’acido nucleico in soluzione in particolare per doppi filamenti formati da  $N_b = 12$  basi azotate. Tale fenomeno porta ad una variazione fino al 70% della concentrazione critica a cui il sistema transisce in una fase LC e non può essere spiegato dalla sola presenza del rapporto tra le dimensioni geometriche  $X_0$  del DNA.

La spiegazione di questo fenomeno si deve quindi ricercare nella conformazione dinamica che il doppio filamento assume quando immerso in soluzione in funzione della sua composizione

chimica. Abbiamo pensato che fosse ragionevole attribuire un ruolo di cruciale importanza della struttura media di equilibrio dell'acido nucleico immerso in soluzione, cercando di spiegare il fenomeno di nematizzazione a diverse concentrazioni mediante sole proprietà di piegatura del doppio filamento di DNA corto rispetto al suo solco maggiore. Per valutare il ruolo giocato dal “bending” nella transizione di fase nematica in modo dipendente dalla sequenza considerata ci siamo proposti di suddividere il nostro lavoro in due parti principali.

Nella prima, grazie all'ausilio della dinamica molecolare “full atom” abbiamo osservato e valutato quantitativamente il fenomeno di ripiegamento del segmento corto di DNA sul suo avvitemento maggiore quando immerso in soluzione acquosa a concentrazione ionica  $1M$ , introducendo due osservabili dinamiche che ci consentissero di valutare il grado di piegatura assunto dalla sequenza in soluzione. L'alta concentrazione salina è stata scelta per schermare le interazioni elettrostatiche ed è in linea con quanto fatto da Nakata [16] nei suoi esperimenti. Mediante alcune analisi preliminari delle simulazioni a singola sequenza fortemente diluita effettuate in box cubici, con concentrazioni di DNA molto inferiori ai valori sperimentali

di nematizzazione, abbiamo compreso il ruolo svolto dagli effetti di taglia finita e come la concentrazione salina influisse sulla velocità di raggiungimento di configurazioni stabili per la molecola, schermando i contributi elettrostatici determinati dall'interazione con la sua stessa immagine. Pur riscontrando la differente piegatura delle diverse sequenze in quasi tutte le simulazioni, abbiamo ritenuto opportuno analizzare dinamicamente un sistema più realistico mediante l'introduzione di 18 acidi nucleici interagenti nello stesso box, replicando computazionalmente gli esperimenti di Nakata e Zanchetta [16, 17], fissando pressione e temperatura ai valori tipici di un esperimento reale. Idealmente abbiamo disidratato il sistema che prima, nelle simulazioni a singolo DNA, era molto diluito: ora le concentrazioni di acido nucleico in soluzione sono rese comparabili con quelle sperimentali di nematizzazione, attestandosi a valori di  $c_{18DNA} \simeq 300mg/ml$ .

Grazie all'osservabile di “bending” assiale, da noi ideato prendendo spunto dagli studi conformazionali del lavoro di Dickerson [44], e attraverso un osservabile puramente geometrico che abbiamo definito osservando le caratteristiche posizionali degli atomi di fosforo nella struttura del DNA, siamo stati in grado di riscontrare quanto previsto in linea teorica. La sequenza ACC

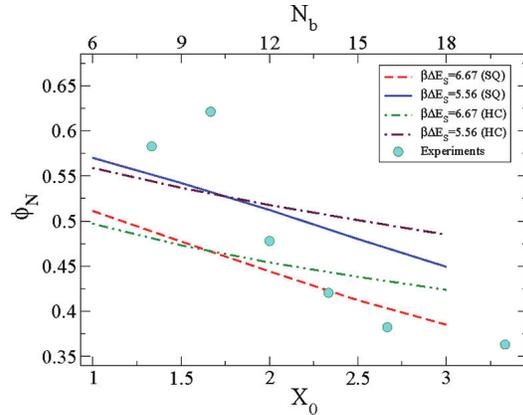


Figura 13.11: Andamento ottenuto nel lavoro di De Michele [19] della  $\varphi_N$  in funzione dell'“aspect ratio”  $X_0$  e valori sperimentali ottenuti. Con le energie usate nelle nostre simulazioni MC ci troviamo al di sotto di tali valori.

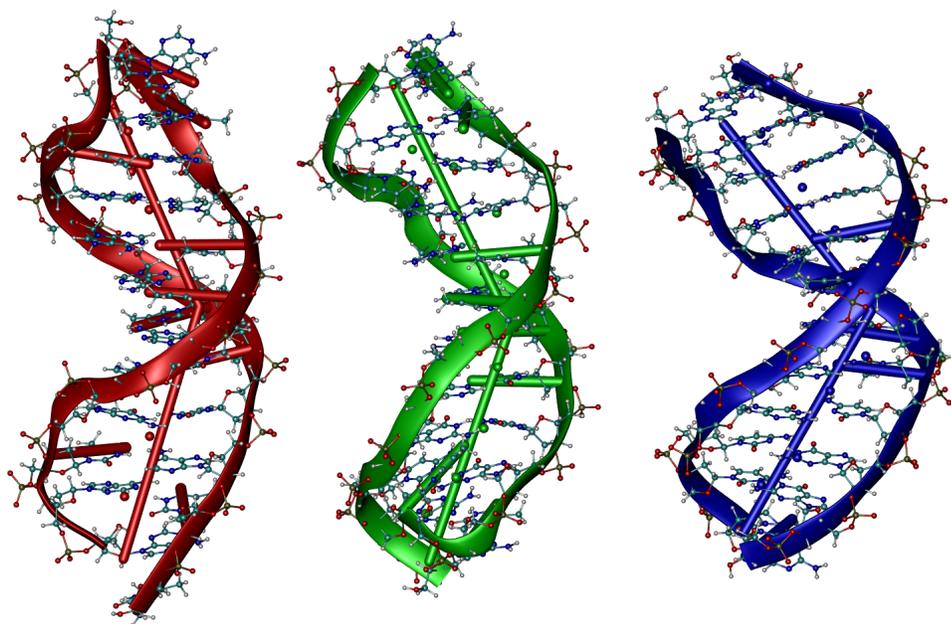


Figura 13.12: “Snapshot” delle configurazioni strutturali più probabili delle molecole di DNA considerate. Usando lo schema RGB la figura rossa rappresenta la sequenza AAT, la verde la AAC e la blu è la più piegata ACC. Il rendering è stato effettuato con il software POV-Ray.

che secondo le misure sperimentali [16] transisce a concentrazioni più alte ( $c_N^{ACC} = 850\text{mg/ml}$ ) è risultata sempre quella maggiormente piegata delle tre durante tutta la durata delle nostre simulazioni (400ns), mentre le altre due sono risultate affette da fenomeni di piegatura minore con entrambi gli osservabili introdotti nel nostro lavoro. Questo ci ha permesso di notare una correlazione stretta tra le concentrazioni di nematizzazione sperimentali e il “bending” delle strutture. Si potrebbe pensare che dato che la struttura si chiude su se stessa sarà anche mediamente più densa e durante la nematizzazione occorreranno concentrazioni maggiori per la transizione, ma in realtà questa spiegazione è valida solo in parte poiché la fase nematica che si forma durante l’assemblamento di oggetti piegati risentirà entropicamente dell’orientazione con cui i singoli doppi filamenti si aggregano tra loro, giocando un ruolo determinante nella termodinamica del sistema.

Una volta ottenute le stime per i valori del “bending” assunto dalle strutture degli acidi nucleici, abbiamo poi pensato di modificare il modello “coarse grained” a cilindro rigido introdotto nei lavori [25, 19] mediante la piegatura dello stesso rispetto all’asse principale. La struttura che abbiamo deciso di creare è stata ottenuta mediante la sovrapposizione di due cilindri disallineati con assi piegati ad un angolo di “bending”  $\Theta$ , “hard bent cylinders” HBC, e rappresenta il singolo “building block” che ci consentirà di modellizzare la struttura piegata assunta dall’acido nucleico in soluzione. Grazie a questa nuova geometria siamo stati in grado

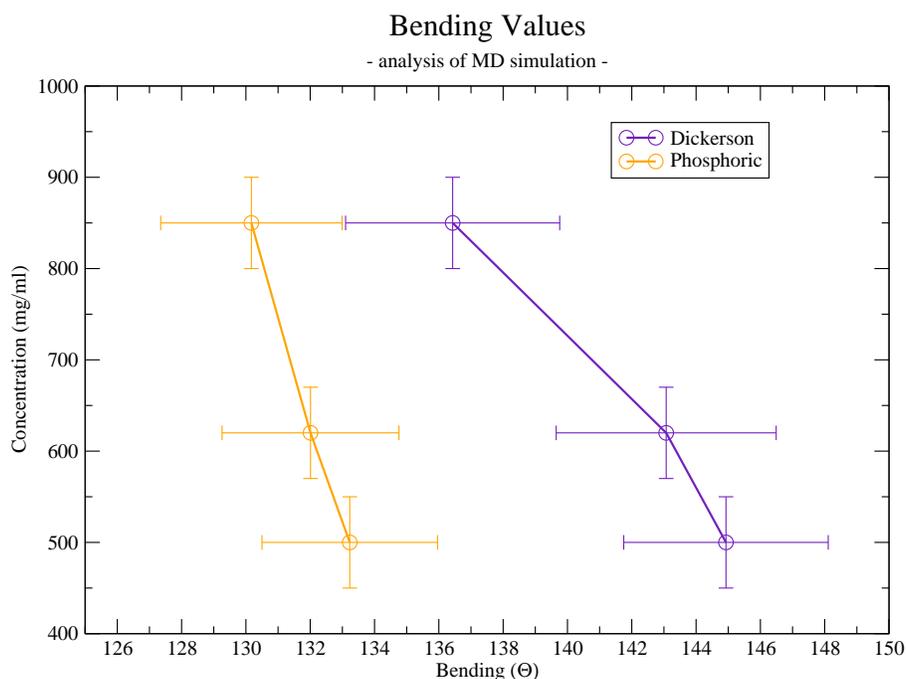


Figura 13.13: Dipendenza delle concentrazioni critiche di nematizzazione ottenute sperimentalmente da Nakata e Zanchetta [16] in funzione del “bending” calcolato con le due osservabili ideate da noi per l’analisi dati delle traiettorie MD a 18 DNA.

di studiare la termodinamica di un sistema composto da un gran numero di DNA piegati interagenti tra loro, permettendo l’analisi della termodinamica del sistema e consentendoci di raggiungere la nematizzazione dovuta alla “self assembly”.

Contestualmente a quanto ottenuto dalle osservabili dinamiche ottenute dalle simulazioni MD, abbiamo scelto di esplorare un “range” di valori di piegatura di circa  $\sim 25^\circ$ , effettuando simulazioni Monte Carlo NPT per sistemi a “bending” variabile nell’intervallo  $\Theta = 134^\circ \div 160^\circ$ , caratterizzandone di volta in volta il diagramma di fase. L’osservazione combinata della discontinuità nel diagramma di fase e del corrispettivo salto del parametro d’ordine di nematizzazione, ci ha permesso di ricavare le “volume fraction” che delimitano la regione di coesistenza delle fasi isotropica e nematica al variare della piegatura  $\Theta$ . Anche in questo caso i risultati ottenuti si comportano come previsto: aumentando il “bending” infatti, aumenta la densità necessaria al sistema per consentire la transizione di fase nonostante il fatto che tali valori rappresentassero delle stime inferiori per colpa del mancato raggiungimento della fase nematica in alcuni dei sistemi considerati. Inoltre grazie all’analisi visiva delle configurazioni dei sistemi più piegati abbiamo osservato l’effetto entropico dovuto alla rottura di simmetria di rotazione azimutale rispetto a strutture cilindriche non piegate: la formazione di aggregati

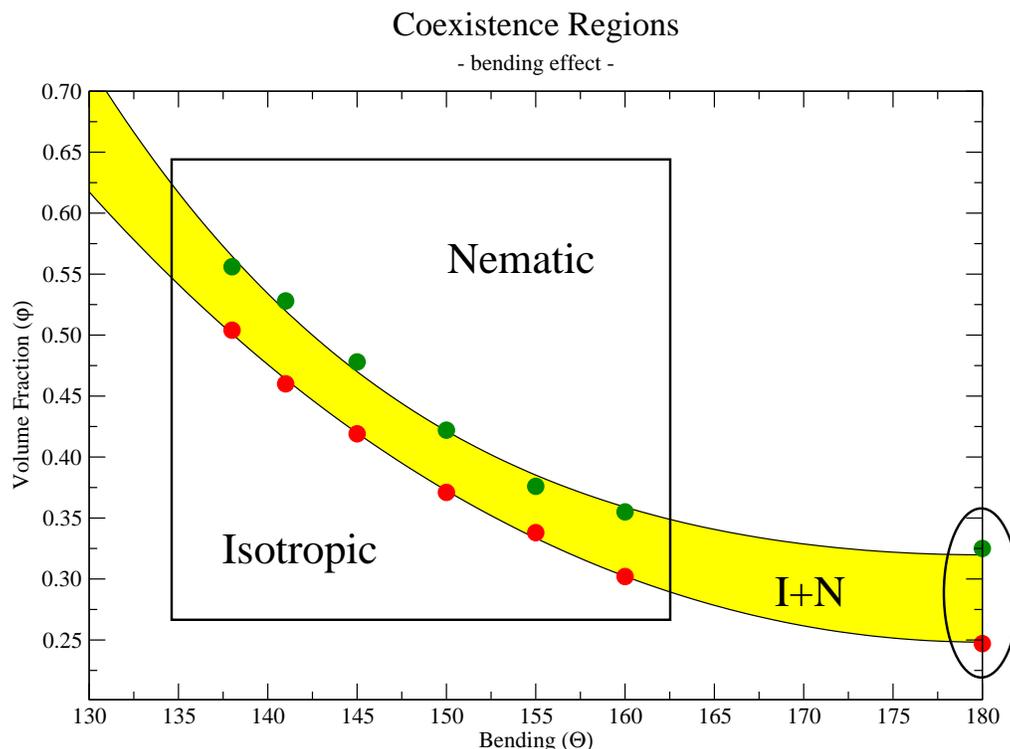


Figura 13.14: Definizione delle regioni di coesistenza delle fasi isotropica e nematica dei doppi cilindri piegati in funzione del grado di “bending” degli oggetti HBC. Ricordiamo che tale curva rappresenta una situazione in cui l’energia di “stacking” tra le basi è pari a  $\beta\Delta E_S = 8.33$  a causa del settaggio della temperatura a  $T = 0.12$ . Le curve che delimitano la regione sono puramente a scopo indicativo e sono state ottenute mediante un fit dei dati sperimentali con una funzione polinomiale quadratica.

a forma di anello, chiusi su loro stessi, diventa via via più probabile con l’aumentare del grado di piegatura, rendendo la transizione di fase nematica sempre più difficile richiedendo quindi molti più passi MC per raggiungere l’equilibratura del sistema e contribuendo all’aumento della “volume fraction” critica di nematizzazione  $\varphi_N$ .

Non solo abbiamo ottenuto tali stime per le “volume fraction” e previsto l’andamento della regione di coesistenza tra le fasi come illustrato in figura 13.14, ma siamo riusciti anche ad osservare l’intervallo di variazione della densità aspettato. Negli esperimenti di Nakata [16] venivano registrate variazioni della concentrazione critica fino al 70% tra le sequenze costituite da 12 basi azotate prese in esame, del tutto in linea con quanto ottenuto da noi attraverso la simulazione di oggetti HBC in cui in  $22^\circ$  apprezziamo una variazione della concentrazione del 57%. Inoltre abbiamo notato che i valori delle “volume fraction” nematiche, ottenute dalle simulazioni ad alti gradi di piegatura, rappresentano delle sottostime della reale situazione

termodinamica poiché non siamo stati in grado di equilibrare tali sistemi a causa del gran numero di passi MC richiesti<sup>4</sup>. Ovviamente una corretta equilibratura avrebbe portato ad un ulteriore aumento delle  $\varphi_N$  che abbiamo pensato di stimare osservando l'andamento del parametro d'ordine in funzione degli step Monte Carlo eseguiti, osservando quali tra i sistemi considerati stesse transendo e riportando la sua “volume fraction”.

Avendo portato alla luce un quadro così vasto di fattori coinvolti nei processi di “self assembly” di doppie eliche corte di DNA molto ancora possiamo fare per comprendere sempre più nel dettaglio la situazione. Per prima cosa in un prossimo futuro ci proponiamo di considerare energie differenti, poiché la scelta di  $\beta\Delta E_S = 8.33$  porta a considerare energie di “stacking” troppo grandi. Come anche mostrato dalla figura 13.11 riportata nel lavoro di De Michele [19] sembra che siano necessari valori più bassi per tali energie, il che porta ad una ulteriore sottostima delle densità di transizione da noi ottenute. Ma in questo modo possiamo pensare di aver determinato una stima superiore per le energie di “stacking” che, considerando il sistema a temperatura ambiente, risulteranno:

$$\Delta E_S < 3.45 \cdot 10^{-20} J = 8.33 k_B T \quad G_{bind} > -2.4 \frac{kcal}{mol} \quad (13.9)$$

dove  $G_{bind}$  è l'energia di stacking espressa dalle unità di misura scelte da Maffeo [21].

Visto il grande interesse nell'argomento, sia per comprendere nel dettaglio la natura delle interazioni e sia per le possibilità concretamente offerte nella creazione di nuovi materiali, ci sembra opportuno sviluppare alcuni aspetti futuri estremamente importanti. In primo luogo sarà necessaria una individuazione più accurata della regione di coesistenza, estendendo ulteriormente le simulazioni Monte Carlo oppure servendoci di algoritmi più efficienti come il “Successive Umbrella Sampling”. Inoltre occorrerà irrobustire la teoria alla base di tale processo, introducendo i contributi entropici azimutali introdotti dalla piegatura della struttura cilindrica finora assenti per ragioni di simmetria. Anche lo studio di sequenze con un diverso numero di “base pairs” sarà cruciale, per mostrare che effettivamente il comportamento studiato si manifesti con tutte le sequenze studiate da Nakata, Zanchetta e Bellini [17, 16] ed andrà sicuramente aggiustata l'energia di “stacking” cercando di riallinearci ai risultati sperimentali ottenuti.

In secondo luogo ci proponiamo di ideare un modello che riscalda le dimensioni del doppio cilindro, mantenendo il volume costante pari a quello del cilindro rettilineo corrispondente, mentre occorrerebbe caratterizzare meglio la definizione delle osservabili di piegatura nella prima parte del testo. A questo scopo potrebbe essere utile effettuare nuovamente delle simulazioni bloccando tra loro le ultime basi di ciascuna sequenza mediante l'introduzione di una forza elastica tra la congiungente dei baricentri degli ultimi nucleotidi, come fatto nelle simu-

---

<sup>4</sup>Da quanto emerso sembra che servano almeno  $10^7$  passi MC per visualizzare il salto nella transizione di fase delle sequenze più piegate.

lazioni di Maffeo [21], per scongiurare effetti di “fraying” delle ultime basi, e successivamente ricalcolare le osservabili. Inoltre potrebbe essere interessante fare un “fit” della struttura dinamica con una struttura a doppio cilindro per mappare direttamente la conformazione molecolare con una struttura HBC. In questa sede ci siamo dovuti limitare all’analisi appena discussa, tentando di amalgamare un gran numero di informazioni a nostra disposizione pur cercando di mantenere un periodico confronto con i più recenti lavori di ricerca fatti finora.

Dopo aver ormai evidenziato dettagliatamente che sequenze corte costituite da una diversa disposizione di basi azotate piegano in modo differente tra loro, quantificando tali effetti attraverso l’analisi di osservabili dinamiche, ed avendo esplorato dal punto di vista statistico la termodinamica di oggetti cilindrici piegati osservando che le densità di nematizzazione aumenta in modo non lineare con il “bending” dei singoli costituenti, possiamo concludere la nostra analisi affermando di aver trovato una nuova grandezza fisica che governa le interazioni nella “soft matter”. Una caratteristica prettamente geometrica come il “bending”, ma come pure lo era il volume escluso della teoria OBF, gioca un ruolo di grande rilievo nel determinare l’equazione di stato di sistemi formati da DNA corti e la comprensione sempre più accurata di tale fenomeno fornirà ai ricercatori un nuovo strumento per controllare le reazioni di “self assembly” e di certo impiego nella progettazione di materiali del futuro.

Ancora una volta, nella fisica della materia soffice, la rottura di simmetria in un sistema spalanca nuove porte nello studio delle interazioni tra le sue stesse costituenti, fornendo delle spiegazioni sempre più interessanti nella comprensione di fenomeni fisici celati dietro comportamenti sorprendenti.

# Ringraziamenti

Ed eccomi quì, di fretta ed all'ultimo secondo a scrivere i ringraziamenti che come nelle migliori tesi non si sa mai da dove iniziare.

Mi sembra opportuno intanto cominciare con il mio relatore, Cristiano, una persona disponibilissima che, anche se lui non lo sa, mi ha fatto riscoprire la mia sopita curiosità per la fisica nonché realizzare il mio sogno da bambino di usare un “grosso” supercomputer. Ed un grazie va anche ad Anna, che con la sua pazienza mi ha orientato all'interno del così vasto ed intricato mondo delle simulazioni.

Un grazie enorme anche a voi, Diego, Laura, Elisa e Marco meravigliosi esemplari di fisici col “fisico” ma soprattutto amici che, seppur scritti in ordine casuale e non certo per importanza, risceglierei altre  $N$  volte come compagni di avventure in questa lunga e strana avventura che, senza di voi, ho i miei dubbi che sarebbe continuata...

E poi gli amici quelli storici. Quelli che ci sono stati sempre nel bene e nel male. A Gianluca per il quale le parole non servono a spiegargli quanto gli devo, lo sa già. Ad Andrea per le meglio cazzate, ora anche in versione infermiere, che fa un po' a cazzotti con la sua immagine precedente. A Matteo, che mi sono ricordato ora come si chiama ma che non mi sono scordato mai di quanto mi sia stato d'aiuto. A Emiliano con cui ho condiviso tutto e con cui voglio continuare a farlo. A Simone il MacGyver della situazione che sono felice abbia ritrovato la grinta di una volta. A Daniele “scritto co' la gn” che mi ha insegnato come ci si diverte e si balla davvero. A Jacopo a cui auguro tutto il bene del mondo, sei forte Jà. A Francesco a cui sono debitore delle meglio risate della mia vita. A Moreno per la sua purezza d'animo ed a Francesco per le risate indimenticabili ed a cui auguro ancora le meglio sbronze.

Non mi scordo mica delle consorti! A Rita per i limoni e le mozzarelle di bufala più buone del mondo, a Luigina per il miglior sorriso sulla faccia della terra, a Rachele per avermi fatto l'onore di essere il suo autista nel giorno più importante della sua vita, a Fabiana che mi ha incastrato su discorsi della moda ma che non mi incastrerà mai con il cinema e a Noemi per la sua risata troppo contagiosa. Ma anche ad Elisa, la mamma più bella di tutte, e alla sua Aurora a cui auguro ogni bene.

E a Federica. Per essere stata sempre al tuo posto. Per aver avuto la grinta di affrontare con me anche le situazioni più strane e difficili. Per avermi sempre sopportato ed avermi

insegnato che significa arrabbiarsi e ridere allo stesso tempo. Ti voglio bene.

A Daniela, la mia preziosa infermiera personale che senza di lei chissà come facevo, e a Marco da sempre mio “cugino” in tutto e per tutto.

Ad Alex ed Angelo a cui penso sempre e vorrei trovare più spesso.

E alla mia famiglia, tutta ma proprio tutta da nord a sud, che tanto mi ha dato e a cui tanto spero di riuscire a dare.

A mia nonna che, pure se non capisce che non ho fame a tutte le ore, so perfettamente che mi vuole il bene più puro e cristallino del mondo. Ed io ricambio, a modo mio. Ai miei nonni che non ci sono più ma che porto sempre con me in qualsiasi cosa che faccio. Davvero.

A mia madre e mio padre che non mi hanno mai messo confini, due genitori che uno vorrebbe avere anche rinascendo altre mille volte. Grazie di tutto.

E alla mia sorellina, a cui in realtà voglio un bene dell’anima anche se non glielo ho mai detto e non glielo dirò mai. Ops l’ho fatto... Non so come farò se mai dovessi separarmi da lei. Ah quasi mi scordavo! La poesia all’inizio è per te e l’ha usata quello brutto in prima pagina quando si inventò letteralmente la materia che ho studiato, in italiano suona così:

*“Abbi divertimento sulla terra e sul mare*

*Infelice è il diventare famoso!*

*Ricchezze, onori, false illusioni di questo mondo,*

*Tutto non è che bolle di sapone”.*

Ancora Grazie a tutti voi e a chi, per la mente offuscata e la fretta del momento, non ho citato, colori vivaci di una vita altrimenti in bianco e nero.

# Bibliografia

- [1] P. G. de Gennes, “*Soft Matter*”, Nobel Lecture, December 9 (1991);
- [2] I. Hamley, “*Introduction to Soft Matter*”, Wiley & Sons, New York (2007);
- [3] R. A. L. Jones, “*Soft Condensed Matter*”, Oxford University Press (2002);
- [4] M. Kleman, O. D. Lavrentovich, “*Soft Matter Physics - An Introduction*”, Springer-Verlag New York Inc. (2003);
- [5] D. Frankel and B. Smit, “*Understanding Molecular Simulation*”, Academic Press, San Diego, California, 2002;
- [6] A. Satoh, “*Introduction to Practice of Molecular Simulation*”, (2011) Elsevier Inc.;
- [7] Y. Li, Y. D. Tseng, S. Y. Kwon, L. D’Espaux, J. S. Bunch, P. L. McEuen and D. Luo, “*Controlled assembly of dendrimer-like DNA*”, Nature Materials Vol. 3, January (2004);
- [8] S. H. Um, J. B. Lee, N. Park, S. Y. Kwon, C. C. Umbach and D. Luo, “*Enzyme-catalysed assembly of DNA hydrogel*”, Nature Materials Vol. 5, October (2006);
- [9] P. W. K. Rothmund, “*Folding DNA to Create Nanoscale Shapes and Patterns*”, Nature Vol. 440, 16 March (2006);
- [10] T. Bellini, R. Cerbino, G. Zanchetta, “*DNA-Based Soft Phases*”, Top. Curr. Chem. 318, 225-280 (2012);
- [11] C. A. Mirkin, R. L. Letsinger, R. C. Mucic, J. J. Storhoff, “*A DNA-based method for rationally assembling nanoparticles into macroscopic materials*”, Nature, Vol. 382, 15 August (1996).
- [12] S.C. Glotzer and J.A. Anderson, “*Nanoparticle Assembly: Made to Order*”, Nature Materials 9 (11), 885-887 (2010);
- [13] N. C. Seeman, “*DNA in a Material World*”, Nature 421, 33-37 (2003);

- [14] L. Onsager, “*The Effects of Shape on the Interaction of Colloidal Particles*”, Ann NY Acad Sci 51:627–659 (1949);
- [15] P. Bolhuis, D. J. Frenkel, “*Tracing the Phase Boundaries of Hard Spherocylinders*”, Chem. Phys. 1997, 106, 666–687.
- [16] M. Nakata, G. Zanchetta, B. D. Chapman, C. D. Jones, J. O. Cross, R. Pindak, T. Bellini, N. A. Clark, “*End-to-End Stacking and Liquid Crystal Condensation of 6-to 20-Base Pair DNA Duplexes*”, Science 318, 1276 (2007);
- [17] G. Zanchetta, F. Giavazzia, M. Nakata, M. Buscaglia, R. Cerbino, N. A. Clark, and T. Bellini, “*Right-handed Double-Helix ultrashort DNA yields Chiral Nematic Phases with both Right- and Left-Handed Director Twist*”, PNAS, October 12 (2010), vol. 107, no. 41, 17497–17502;
- [18] C. De Michele, R. Schilling and F. Sciortino, “*Dynamics of Uniaxial Hard Ellipsoids*”, Physical Review Letters 98, 265702 (2007);
- [19] C. D. Michele, T. Bellini, and F. Sciortino, “*Self-Assembly of Bifunctional Patchy Particles with Anisotropic Shape into Polymers Chains: Theory, Simulations, and Experiments*”, Macromolecules (2012) 45, 1090–1106;
- [20] C. D. Michele, L. Rovigatti, T. Bellini and F. Sciortino, “*Self-assembly of short DNA duplexes: from a coarse-grained model to experiments through a theoretical link*”, Soft Matter (2012) 8, 8388;
- [21] C. Maffeo, B. Luan and A. Aksimentiev, “*End-to-end Attraction of Duplex DNA*”, Nucleic Acids Research (2012) 1–10;
- [22] R. Dahm, “*Discovering DNA: Friedrich Miescher and the early years of nucleic acid research*”, Hum Genet. (2008) Jan, 122(6):565-81;
- [23] R. Dahm, “*Friedrich Miescher and the discovery of DNA*”, Developmental Biology 278 (2005) 274 – 288;
- [24] J.D. Watson, F.H.C. Crick, “*A Structure for Deoxyribose Nucleic Acid*”, Nature (1953) 171: 737–738;
- [25] T. Kuriabova, M. D. Betterton and M. A. Glaser, “*Linear aggregation and liquid-crystalline order: comparison of Monte Carlo simulation and analytic theory*”, arXiv:1007.4530 [cond-mat.soft] (2010);
- [26] P. J. Hagerman, “*Flexibility of DNA*”, Ann. Rev. Biophys. Biophys. Chem. (1988). 17:26-86;

- [27] C. D. Michele, “*Simulating hard rigid bodies*”, Journal of Computational Physics 229 (2010) 3276–3294;
- [28] R. M. Dirks, M. Lin, E. Winfree and N. A. Pierce, “*Paradigms for computational nucleic acid design*”, 1392-1403 Nucleic Acids Research (2004) Vol. 32, No. 4;
- [29] B.H. Bransden, C. J. Joachain, “*Physics of Atoms and Molecules - 2nd Edition*”, Pearson Education (2003);
- [30] J. J. De Pablo, “*Coarse-grained simulations of macromolecules: from DNA to nanocomposites*”, Annu Rev Phys Chem 62:555-74. 2011;
- [31] D. Andreatta, S. Sen, J. L. Pérez Lustres, S. A. Kovalenko, N. P. Ernsting, C. J. Murphy, R. S. Coleman, M. A. Berg, “*Ultrafast dynamics in DNA: "fraying" at the end of the helix*”, J. Am. Chem. Soc. (2006) May 31;128(21):6885-92
- [32] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham, S. De Bolt, D.M. Ferguson, G.L. Seibel, P.A. Kollman. “*AMBER, a package of computer programs applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules*”, Comp. Phys. Comm. (1995) 91, 1-41;
- [33] A. D. Mackerell, J. Wiorkiewicz-Kuczera, M. Karplus , “*An all-atom empirical energy function for the simulation of nucleic acids*”, J. Am. Chem. Soc. (1995) 117, 11946–75;
- [34] M. Orozco, A. Pérez, A. Noy, F. J. Luque, “*Theoretical methods for the simulation of nucleic acids*”, Chem. Soc. Rev. (2003) 32, 350-364;
- [35] A. Morriss-Andrews, J. Rottler and S. S. Plotkin, “*A systematically coarse-grained model for DNA, and its predictions for persistence length, stacking, twist, and chirality*”, J. Chem. Phys. (2010) 132, 035105;
- [36] P. Yakovchuk, E. Protozanova and M. D. Frank-Kamenetskii, “*Base-stacking and base-pairing contributions into thermal stability of the DNA double helix*”, 564–574 Nucleic Acids Research, 2006, Vol. 34, No. 2;
- [37] T.E. Ouldridge, I.G. Johnston, A.A. Louis and J.P.K. Doye, “*The self-assembly of DNA Holliday junctions studied with a minimal model*”, J. Chem. Phys (2009) 130, 065101;
- [38] W. K. Olson, M. Esguerra, Y. Xin, X. J. Lu, “*New information content in RNA base pairing deduced from quantitative analysis of high-resolution structures*”, 177-186 Nucleic Acid Structure, 2009, Vol. 47, No.3;

- [39] K. Drukker, G. Wu and G. C. Schatz, “*Model simulations of DNA denaturation dynamics*”, J. Chem. Phys. (2001) 114, 579;
- [40] M. Kenward and K. D. Dorfman, “*Brownian dynamics simulations of single-stranded DNA hairpins*”, J. Chem. Phys. (2009) 130, 095101;
- [41] E. J. Sambriski, V. Ortiz and J. J. de Pablo, “*Sequence effects in the melting and renaturation of short DNA oligonucleotides: structure and mechanistic pathways*”, J. Phys.: Condens. Matter (2009) 21 034105;
- [42] J. Largo, F. W. Starr and F. Sciortino, “*Self-Assembling DNA Dendrimers: A Numerical Study*”, Langmuir 23, 5896-5905 (2007);
- [43] H. H. Loeffler, M. D. Winn, “*Large biomolecular simulation on HPC platforms III. AMBER, CHARMM, GROMACS, LAMMPS and NAMD*”;
- [44] R. E. Dickerson, “*DNA bending: the prelevance of kinkiness and the virtues of normality*”, 1906-1926 Nucleic Acids Research (1998) Vol. 26, No. 8;
- [45] J. Wang, P. Cieplak, P. A. Kollman, “*How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules?*”, Journal of Computational Chemistry (2000), Vol. 21, No. 12, 1049–1074;
- [46] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, P. Kollman, “*A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations*”, Journal of Computational Chemistry (2003), Vol. 24, No. 16;
- [47] P. Mark and L. Nilsson, “*Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K*”, J. Phys. Chem. A (2001), 105, 9954-9960;
- [48] A. Y. Toukmaji, J. A. Board Jr., “*Ewald summation techniques in perspective: a survey*”, Computer Physics Communications 95 (1996) 73-92;
- [49] T. Darden, L. Pedersen, A. Toukmaji, M. Crowley, T. Cheatham, “*Particle-Mesh based methods for fast Ewald summation in molecular dynamics simulations*”, unpublished;
- [50] C. J. Fennell and J. D. Gezelter, “*Is the Ewald summation still necessary? Pairwise alternatives to the accepted standard for long-range electrostatics*”, the Journal of Chemical Physics, 124, 234104 (2006);
- [51] C. Kittel, “*Introduction to Solid State Physics*”, Wiley, (1971) , New York;
- [52] J.P Hansel, I. R. McDonald, “*Theory of Simple Liquids*”, Elsevier Science, 2006;

- [53] S. Neidle, "*Principles of Nucleic Acid Structure*", Academic Press (2008) San Diego, California;
- [54] F. , N. Špačková, M. Moakher, P. Enkhbayar and J. Šponer, "*A measure of bending in nucleic acids structures applied to A-tract DNA*", 3414-3422 *Nucleic Acids Research* (2010) Vol. 36, No. 10;
- [55] C. Maffeo, B. Luan and A. Aksimentiev, "*Supporting Information: End-to-end Attraction of Duplex DNA*", August 26, 2011;
- [56] J. Hizver, H. Rozenberg, F. Frolow, D. Rabinovich, and Z. Shakked, "*DNA bending by an adenine-thymine tract and its role in gene regulation*", *PNAS* July 17 (2001) vol. 98, no. 15;
- [57] P. van Der Schoot, M. E. Cates, "The Isotropic-to-Nematic Transition in Semi-Flexible Micellar Solutions", *Europhys. Lett.*, (25) (7), pp. 515-520 (1994);
- [58] A. R. Khokhlov, A. N. Semenov, "*Liquid-Crystalline Ordering in the Solution of Partially Flexible Macromolecules*", *Physica A* (1982) vol.112, p.605;
- [59] Odijk, "*Theory of Lyotropic Polymer Liquid Crystals* ", *T. Macromolecules* (1986) 19, 2313;
- [60] P. G. de Gennes, J. Prost. "*The Physics of Liquid Crystals*", Clarendon Press (1995) Oxford, second edition.