

UNIVERSITÀ DEGLI STUDI DI PISA
FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E
NATURALI

CORSO DI LAUREA IN FISICA



TESI DI LAUREA MAGISTRALE

**Simulazioni di dinamica
molecolare classica con modelli
a bassa risoluzione della proteasi
di HIV-1**

Relatore
Valentina Tozzini

Candidato
Marco Galimberti

ANNO ACCADEMICO 2012/2013

Indice

1	Introduzione e motivazione del lavoro	3
2	Il virus HIV	5
2.1	La causa dell'AIDS	5
2.2	Struttura del virus	6
2.3	La poliproteina Gag	9
2.4	Il ciclo virale	11
2.5	Terapie virali	11
2.6	La proteasi di HIV	12
2.6.1	La reazione di idrolisi	14
2.6.2	Sequenze tagliate dalla proteasi dell'HIV	17
2.6.3	L'ipotesi dell'involuppo del substrato	18
2.6.4	Conclusioni	21
3	Dinamica molecolare classica con modelli a bassa risoluzione	23
3.1	Dinamica molecolare classica	23
3.1.1	Algoritmi di integrazione	23
3.1.2	Termostati	25
3.1.3	Dinamica di Langevin e browniana	26
3.2	Campi di forze atomistici	26
3.3	Campi di forze a bassa risoluzione	27
3.3.1	Modelli a rete elastica	28
3.3.2	Modelli $G\bar{o}$	30
3.3.3	Inversione di Boltzmann	31
3.4	Studi di simulazione della dinamica della proteasi	32
3.5	Studio di simulazione sulla poliproteina Gag	41
3.6	Sommario e motivazione del lavoro di tesi	42
4	Interazione tra proteasi di HIV-1 e poliproteina Gag: ottimizzazione dei campi di forze e simulazioni	45
4.1	Scelta del software DL_POLY	45
4.2	Il campo della proteasi	46
4.2.1	La dinamica dei flap	47
4.2.2	Proteasi e substrato	52
4.3	Ottimizzazione del campo di forze di Gag	60

4.3.1	Dati NMR	60
4.3.2	Scelta strutture	61
4.3.3	Costruzione del campo di forze di Gag	63
4.3.4	Dinamica di Gag da sola e miglioramento del campo di forze	64
4.4	Simulazioni della proteasi con Gag	66
4.4.1	Campo di forze dell'interazione Gag-Proteasi	66
4.4.2	Simulazioni dell'interazione Pr-Gag	66
4.5	Conclusioni	72
5	Conclusioni e possibili sviluppi	75
A	Tabella amminoacidi	77
B	Risultati dei lavori sulla predizione delle sequenze idrolizzate dalla proteasi	78
C	DL_POLY	81
D	Parametri del campo di forze	84
E	Interazioni non bonded locali Pr-substrato	86
F	Tabelle valori RMSD	88
G	Valori degli angoli e dei diedri delle strutture NMR	92
H	Valori degli angoli di MA-CA_{NTD}	94
I	Distribuzioni degli angoli del linker	95
J	Distribuzioni dei diedri del linker	101
K	Grafici delle simulazioni dell'interazione Pr-Gag	107
	Bibliografia	122

Capitolo 1

Introduzione e motivazione del lavoro

La proteasi del virus HIV-1 è una proteasi aspartica attiva come un omodimero simmetrico, formata cioè da due monomeri uguali aventi una lunghezza di 99 amminoacidi. Questo enzima svolge un ruolo fondamentale nella fase finale del processo di maturazione del virione di HIV tagliando tramite idrolisi del legame peptidico, le poliproteine Gag e Gag-Pol in elementi funzionali costitutivi del virione infettivo. L'inibizione di questo processo produce virioni non infettivi, ed è tutt'ora considerata una delle migliori strategie contro l'AIDS.

In questo lavoro di Tesi si studia la dinamica molecolare del processo di azione della proteasi sulla Gag, allo scopo di individuarne possibili strategie inibitorie. Il complesso Gag-proteasi è relativamente grande su scala atomica, comprendendo due proteine ciascuna composta da diversi domini. La Gag, infatti, pure essendo costituita da un'unica catena polipeptidica, è strutturata in domini relativamente rigidi, collegati da parti destrutturate e flessibili, di solito contenenti i siti di taglio che la proteasi riconosce.

Inoltre, è noto da dati sperimentali di varia natura che il processo di apertura del sito attivo della proteasi ed aggancio alla Gag avviene su scale dell'ordine del micro-millisecondo. La combinazione di grandi dimensioni spaziali e scale temporali lunghe rende il problema difficilmente affrontabile con metodi di dinamica molecolare classica al livello atomistico. Si è pertanto usato un modello a bassa risoluzione, precisamente al livello di un singolo centro interattivo per amminoacido. Questo riduce notevolmente il numero di gradi di libertà interni rendendo possibile la simulazione del processo con risorse computazionali modeste.

La proteasi idrolizza la Gag in diversi punti, generando così diverse proteine del nuovo virione. Questo lavoro si focalizza su uno specifico sito di taglio, quello tra i domini MA e CA, che genera una proteina costitutiva del capsido del nuovo virione, ed è quindi un passo essenziale della maturazione della particella virale. La scelta dello specifico sito di taglio è stata operata dopo un'analisi statistica dei possibili siti di taglio che è interessante di per sé perché da informazioni sui meccanismi di riconoscimento e specificità della proteasi dell'HIV. Questa analisi contenuta nel prossimo capitolo, che è in generale dedicato alla descrizione del virus HIV e del suo processo di replicazione, evidenzia i motivi per cui risulta difficilmente contrastabile. Inoltre il capitolo introduce e descrive l'enzima proteasi di HIV-1 e la Gag, mostrando come si suppone avvenga il riconoscimento della sequenza peptidica da tagliare.

Nel terzo capitolo viene descritto il metodo di indagine adottato in questo studio. Si descrive la dinamica molecolare classica in generale e specificamente con modelli a bassa

risoluzione ("coarse-grained") e vengono mostrati i risultati ottenuti da precedenti studi di simulazione sulla proteasi di HIV e sulla Gag.

Nel quarto capitolo sono descritti i campi di forze utilizzati per le due proteine e i risultati ottenuti dalle varie simulazioni. Per la proteasi, si è utilizzato un modello precedente esistente in letteratura, che mostrava di riprodurre correttamente la dinamica e la statistica di apertura e chiusura del sito attivo. Per la Gag e per le interazioni proteasi-Gag, invece si è costruito *de novo* un modello compatibile con quello della proteasi. Nel modello della Gag sono stati incorporati i dati strutturali disponibili, sia da cristallografia a raggi X che NMR. Il modello di Gag è stato validato con simulazioni della Gag isolata, che mostrano riprodurre la flessibilità attesa della proteina, grazie all'utilizzo di forme funzionali non standard ed un'accurata parametrizzazione. È stata inclusa nel modello anche la conoscenza a priori sul sito di taglio in modo da rendere selettiva e specifica l'interazione tra Gag e proteasi.

I modelli di Gag e proteasi sono stati poi accoppiati per simulare la dinamica di aggancio della proteasi alla Gag. Questa è la prima simulazione di questo genere in cui l'intero dominio di interesse della Gag viene incluso, rivelando alcuni dettagli molecolari della dinamica di apertura del sito e del riconoscimento della sequenza di taglio. L'analisi delle simulazioni rivela il ruolo della flessibilità della Gag nel processo e dà indicazioni su possibili strategie di inibizione.

Capitolo 2

Il virus HIV

2.1 La causa dell'AIDS

Nell'anno 1983 è stato verificato che il virus dell'HIV (Human immunodeficiency virus) è il responsabile della *sindrome di immunodeficienza acquisita* (AIDS) [1, 2]. Alla fine del 2010 è stato stimato che 34 milioni di persone nel mondo erano infettate da HIV, con un aumento del 17% rispetto al 2001. Questo dato riflette da un lato il grande numero di nuove infezioni di HIV, dall'altro una notevole espansione all'accesso delle terapie, che certamente ha aiutato a ridurre le morti dovute all'AIDS specialmente negli ultimi anni. Il numero di persone che muoiono per cause legate all'AIDS è sceso a circa 1.8 milioni nel 2010, scendendo da un picco di circa 2.2 milioni negli anni 2005 e 2006 (<http://www.unaids.org/>), aumentando l'aspettativa di vita dei sieropositivi e di conseguenza la loro percentuale nella popolazione. Questo a sua volta facilita la diffusione del virus.

HIV è trasmesso tramite contatti sessuali diretti, dal sangue e dai suoi derivati e da una madre infetta al bambino durante la gravidanza, al momento del parto o dopo la nascita tramite l'allattamento [3]. Non ci sono prove sul fatto che HIV possa essere trasmesso attraverso normali contatti sociali o da punture di insetti come le zanzare.

Ci sono due tipi di HIV, HIV-1 e HIV-2 che discendono dal virus SIV, Simian immunodeficiency virus, infettante le scimmie. HIV-1 probabilmente è stato originato da SIVcpz presente negli scimpanzé meno di 100 anni fa mentre HIV-2, un virus ancora principalmente confinato nell'Africa occidentale, è stato probabilmente originato da SIVsm presente nei cercocebi, un'altra specie di scimmie. HIV-1 e HIV-2 a loro volta si dividono in gruppi e successivamente in sottotipi. HIV-1 comprende quattro differenti lignaggi, chiamati gruppi M, N, O e P ognuno dei quali deriva da un evento indipendente di trasmissione, e di cui il gruppo M è stato il primo ad essere scoperto e rappresenta la forma pandemica di HIV-1. HIV è un retrovirus, caratterizzato cioè da genoma a base di RNA (acido ribonucleico), e che necessita quindi di una proteina specifica (trascrittasi inversa) per trascrivere RNA in DNA e usare così il macchinario replicativo della cellula infettata. Questo tipo di virus è generalmente caratterizzato da elevata variabilità, infatti il gruppo M è a sua volta suddiviso in nove sottotipi ($A - D, F - H, J, K$) e nelle forme ricombinanti CRFs che sono un incrocio tra due sottotipi (vedi figura 2.1) [4]. Un retrovirus presenta infatti una mutazione ogni $10kb$ (una mutazione ogni 10^4 basi) ad ogni replicazione del genoma e a causa di questo elevato

tasso di mutazione è difficile da contrastare.

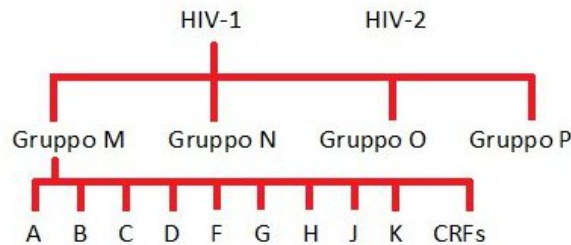


Figura 2.1: Schema della suddivisione di HIV in tipi, gruppi e sottotipi

2.2 Struttura del virus

Il virus dell'HIV appartiene al VI gruppo della classificazione di Baltimore (figura 2.2) [5], che classifica i virus sulla base di proprietà strutturali del genoma (DNA o RNA, senso o antisenso, singolo o doppio filamento) e che determina però anche il numero di passaggi che si devono compiere per trasformare il genoma virale in RNA messaggero compatibile con il macchinario cellulare e utilizzabile per produrre proteine virali e quindi nuove copie del virus. Sono state così ottenute sette classi:

- Classe I: Virus a DNA a doppio filamento.
- Classe II: Virus a DNA a singolo filamento a senso positivo
- Classe III: Virus a RNA a doppio filamento
- Classe IV: Virus a RNA a singolo filamento a senso positivo
- Classe V: Virus a RNA a singolo filamento a senso negativo
- Classe VI: Virus a RNA a singolo filamento a senso positivo con intermedio DNA
- Classe VII: Virus a DNA a doppio filamento con intermedio RNA

Il gruppo VI contiene i retrovirus (retroviridae) a genoma (diploide) a singolo filamento di RNA a polarità positiva (+ssRNA). Questa struttura comporta che il genoma debba prima essere trasformato in DNA a doppia elica, poi integrato nel DNA della cellula, compiti svolti da due specifiche proteine virali, la trascrittasi inversa (RT) e l'integrasi (IN). Una terza proteina virale, la proteasi (PR) interviene nella fase di maturazione. Il genoma e le proteine

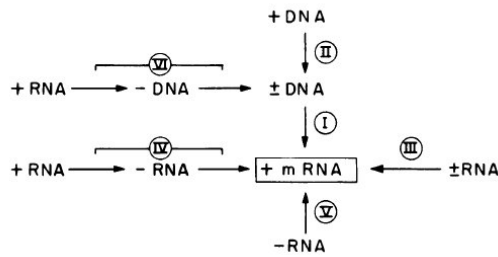


Figura 2.2: Schema originario della classificazione di Baltimore che illustra come i diversi gruppi virali (rappresentati da un numero romano cerchiato) arrivano diversamente alla sintesi dell'RNA messaggero. Per ogni gruppo è mostrato l'acido nucleico di partenza, se DNA o RNA, a senso positivo (+) o negativo (-), a singolo (+ o -) o doppio filamento (\pm), ed il percorso che deve compiere per arrivare a sintetizzare l'RNA messaggero. Successivamente è stato aggiunto un settimo gruppo per distinguere i virus aventi inizialmente un a doppio filamento di DNA ma con un intermedio a RNA [5].

virali sono contenuti nel capsido, struttura formata dall'unione di singole specifiche proteine combinate in modo da formare esagoni e pentagoni come in un grande fullerene tronco-conico.

HIV è del genere lentivirus che significa che è un virus lento, cioè che infetta l'organismo ospite ma i suoi sintomi si manifestano solo dopo un certo periodo di tempo. Inoltre HIV è dotato di un pericapside (detto anche involucro) che è costituito da un doppio strato fosfolipidico al quale sono ancorate glicoproteine virali e che contiene il capsido. Le due glicoproteine (proteine contenenti carboidrati legati covalentemente alla struttura polipeptidica) virali dell'involucro, *gp120* e *gp41*, sono responsabili rispettivamente dell'attacco e della fusione con la membrana cellulare.

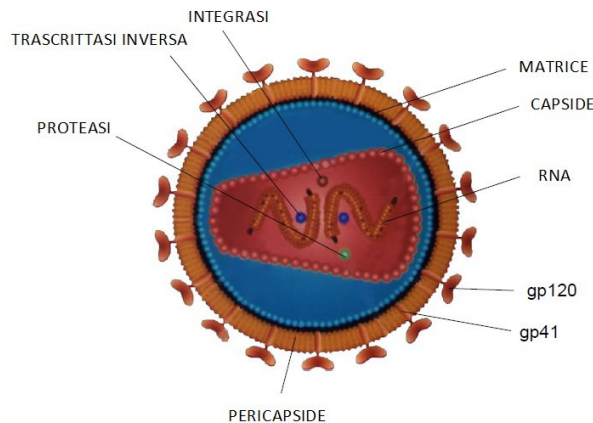


Figura 2.3: Struttura del virione maturo di HIV

HIV colpisce varie cellule del corpo umano, in particolare quelle con recettore *CD4*, una proteina che solitamente agisce nel sistema immunitario, a cui si lega la proteina di superficie *gp120*. Queste interazioni non sono però sufficienti affinché avvenga l'infezione, perché è

necessario che *gp120* si leghi con un'altra proteina presente sulla superficie della cellula bersaglio, chiamata co-recettore, per permettere l'entrata del virus nella cellula. Nel caso di HIV-1 i più comuni co-recettori sono *CCR5* e *CXCR4*. A causa dei legami con il recettore ed il co-recettore, *gp120* subisce un cambiamento conformazionale e permette l'esposizione di *gp41* che riesce così a fondersi con la membrana cellulare permettendo l'ingresso del materiale virale all'interno della cellula.

Da studi di microscopia elettronica le particelle virali, chiamate virioni, hanno le dimensioni comprese tra 80-120 nm di diametro [6].

Il genoma di HIV contiene nove geni fondamentali:

- i tre geni strutturali *gag*, *pol*, *env* che sono tipici di tutti i retrovirus, servono per codificare gli elementi necessari per la replicazione virale
- i geni regolatori *rev*, *tat*
- i geni accessori *nef*, *vpr*, *vpu* e *vif*

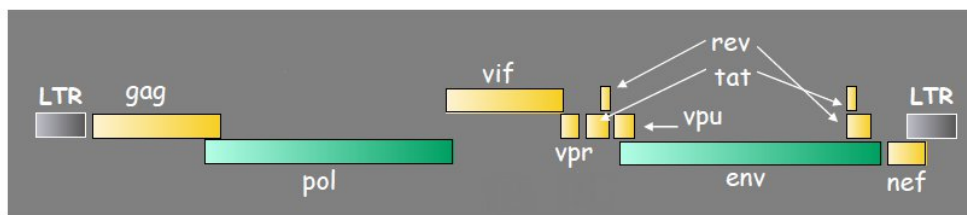


Figura 2.4: Rappresentazione del genoma di HIV

A ciascuna estremità delle catene di RNA virale c'è una serie ripetuta di frequenze, dette unità ripetitive terminali o *Long Terminal Repeat* (LTR) che hanno un ruolo importante nell'integrazione e nella trascrizione dei geni di HIV.

I geni *gag* e *pol* sono inizialmente trascritti in un unico mRNA che porta alla traduzione delle poliproteine Gag e Gag-Pol. La prima verrà prodotta dalla traduzione dell'mRNA dal codone¹ d'inizio fino al primo codone di terminazione, il superamento del quale determina invece l'espressione del gene *pol* [7], il quale codifica per le tre proteine virali a funzione enzimatica, ovvero RT, PR e IN [6].

Durante il processo di maturazione di HIV-1, la proteasi taglia le poliproteine virali Gag e Gag-Pol rendendo possibile l'assemblaggio in un virione infettivo (vedi figura 2.5). Studiare l'enzima proteasi serve quindi a capire meglio come avviene il processo di maturazione del virione ed è utile per cercare di bloccare questo processo mantenendo i virioni non infettivi e quindi contrastando l'espansione del virus HIV-1.

¹Per codone si definisce un tratto di DNA avente la lunghezza di tre nucleotidi. Ogni nucleotide è una molecola formata da una base azotata, uno zucchero e un gruppo fosfato; è l'unità costitutiva fondamentale degli acidi nucleici. Generalmente un codone codifica un singolo amminoacido, ma in alcuni casi può causare l'inizio o la terminazione della traduzione dell'RNA messaggero nella proteina.

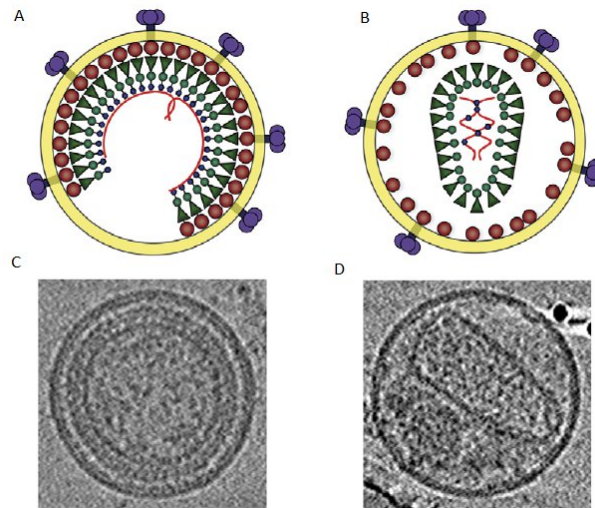


Figura 2.5: Modelli schematici dei virioni di HIV-1 immaturi (A) e maturi (B). Sezione centrale tomografica dei virioni di HIV-1 immaturi (C) e maturi (D) ottenuti dalla criotomografia elettronica [8].

2.3 La poliproteina Gag

Le proteine strutturali di HIV-1 derivano tutte dalla poliproteina Gag, dalla quale è generato un set di proteine e peptidi distanziatori: MA, CA, SP1, NC, SP2, e p6 (vedi figura 2.6). La regione CA è costituita da due domini: il dominio NTD (da *N-terminal domain*), che partecipa all'assemblaggio del nucleo virale (costituito dall'RNA e dalle proteine accessorie), e il dominio CTD (*C-terminal domain*) che partecipa sia all'assemblaggio del virione immaturo sia del nucleo [9].

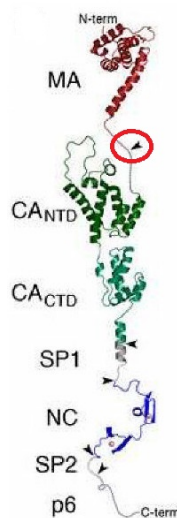


Figura 2.6: Modello strutturale della poliproteina gag, derivato da modelli di domini isolati. Le frecce indicano i siti di taglio. Le linee tratteggiate rappresentano regioni destrutturate (linker). Il sito di taglio evidenziato dal cerchietto rosso, sarà oggetto di studio di questa tesi.

Per ottenere queste nuove proteine, la proteasi taglia quindi la poliproteina Gag in cinque punti, durante il processo chiamato maturazione. Nel virione immaturo le molecole Gag sono estese e orientate radialmente [10]. L'estremità N-terminale di MA è legata allo strato più interno della membrana e la parte C-terminale di MA forma un' α -elica di 25 amminoacidi che dalla testa del dominio MA si estende verso il centro del virione [11]. I domini MA e CA(N) sono separati da una distanza di circa 40 Å e sono collegati da un linker flessibile. La regione C-terminale di MA e la regione CA(N) formano una struttura estesa di circa 70 Å [12].

I vari domini della Gag formano quindi dei gusci quasi sferici nel virione immaturo, che sono profondamente modificati durante il processo di maturazione, dove la proteasi agisce sui siti di taglio permettendo un profondo cambiamento conformazionale.

Il taglio dei siti MA-CA, CA-p2 e p2-NC è fondamentale per la replicazione di HIV e avviene secondo un preciso ordine. Il sito p2-NC è tagliato più velocemente, successivamente viene tagliata la giunzione MA-CA e infine il sito CA-p2.

Queste nuove proteine provenienti dalla poliproteina Gag, si riassemblano per formare gli strati distinti del virione maturo: MA rimane associata con la membrana virale interna (la matrice), NC riveste l'RNA virale (il nucleocapside) e CA assembla nel capsido conico (vedi figura 2.7) ciò che circonda il nucleocapside e gli enzimi ad esso associati (la trascrittasi inversa e l'integrasi) [13].

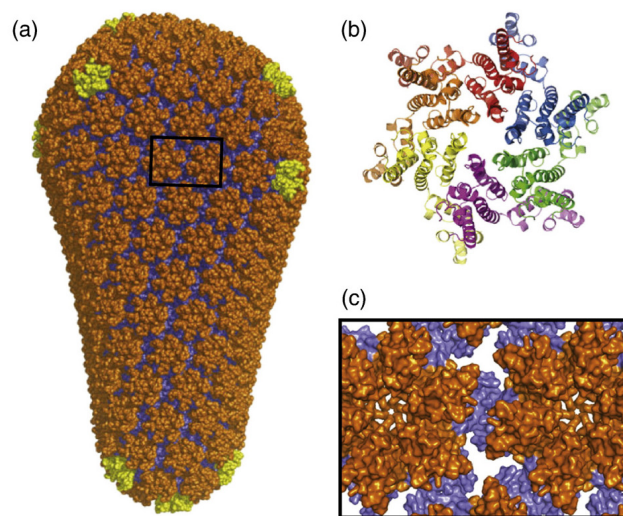


Figura 2.7: Struttura del capsido maturo di HIV. (a) Modello del cono di fullerene. Gli esameri, i pentameri e i dimeri che lo costituiscono, sono colorati rispettivamente in arancione, giallo e blu. (b) Struttura degli esameri formati dal dominio N-terminale di CA. (c) Un ingrandimento della regione selezionata in (a) che illustra le sei interazioni del dominio N-terminale (arancione) e le due interazioni del dominio C-terminale (blu) [14].

Successivamente in questo lavoro di tesi vedremo come la proteasi si aggancia al sito di taglio MA-CA, che quindi svolge un ruolo fondamentale anche in questo lavoro.

2.4 Il ciclo virale

IL ciclo infettivo virale (di cui è data una rappresentazione in figura 2.8) inizia quando il pericapside del virione si fonde con la membrana cellulare per mezzo dell'interazione tra il recettore presente sulla superficie molecolare e le glicoproteine virali sulla superficie del virione. Il materiale virale entra nella cellula grazie al disassemblamento del capsido che libera il genoma e l'RNA del virus viene trascritto in DNA tramite la RT. Viene così prodotto del DNA a doppia elica che viene integrato nel genoma della cellula per mezzo dell'IN. Questo DNA viene successivamente trascritto in mRNA così da permettere la traduzione del DNA virale nelle poli-proteine virali. Il materiale virale necessario alla creazione del virione viene quindi assemblato nel citoplasma in nuovi virioni espulsi dalla cellula tramite il processo di *budding*, in cui i virioni immaturi acquisiscono il pericapside, che deriva dalla membrana citoplasmatica della cellula ospite e per questo motivo è costituito da un doppio strato fosfolipidico. Dopo il processo di budding avviene la maturazione del virione grazie alla PR [6].

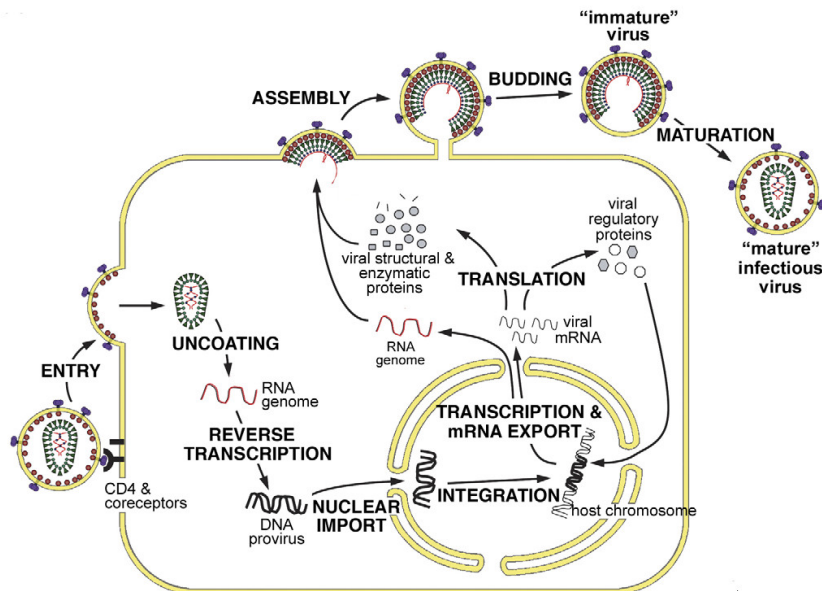


Figura 2.8: Replicazione del ciclo di HIV [15]

2.5 Terapie virali

Lo studio del ciclo virale ha permesso di stabilire i bersagli su cui concentrarsi per attuare le terapie mediche. Attualmente ci sono sette classi di farmaci antiretrovirali (ARV) disponibili nel trattamento dei pazienti infetti da HIV-1:

1. inibitori nucleosidici² della trascrittasi inversa (NRTI)
2. inibitori nucleotidici³ della trascrittasi inversa (NtRTI)
3. inibitori non-nucleosidici della trascrittasi inversa (NNRTI)
4. inibitori della proteasi (Pi)
5. inibitori della fusione (EI)
6. antagonisti dei co-recettori (CRI)
7. inibitori dell'integrasi (INI)

La terapia sui pazienti infetti da HIV-1 si basa sulla combinazione di tre o più farmaci di due o più classi in modo da agire su vari stadi del ciclo di vita del virus. Varie combinazioni di tre o quattro farmaci sono note come HAART o Highly Active Anti Retroviral Therapy [16].

L'inibizione della proteasi di HIV-1 dà come risultato particelle virali immature e non infettive, [17] ed è per questo motivo che la proteasi è uno dei bersagli più importanti delle terapie contro l'HIV [18].

Questo lavoro di tesi si concentra nello studio della proteasi, in quanto la miglior comprensione del funzionamento della proteasi può servire a progettare nuovi e più efficienti inibitori di questo enzima.

2.6 La proteasi di HIV

È stato precedentemente mostrato il ruolo fondamentale che la proteasi di HIV svolge nella fase finale del processo di maturazione andando a tagliare le poliproteine Gag e Gag-Pol per consentire la maturazione del virione.

La proteasi di HIV è una proteasi aspartica attiva come un omodimero simmetrico, in quanto ha due monomeri uguali aventi una lunghezza di 99 amminoacidi. Essi, formati un'elica α e due foglietti β antiparalleli nella struttura secondaria, si legano in conformazione asimmetrica assiale nella proteina funzionale, e contribuiscono al legame con il substrato. Il sito attivo dell'enzima, chiamato anche sito catalitico, dove avviene il taglio del substrato, è formato dai residui 24 – 30 di ciascun monomero e contiene una triade catalitica composta da Asp25, Thr26, Gly27⁴ per ogni monomero. Una caratteristica dell'enzima è che questo suo sito catalitico è coperto da una coppia di "flap", ognuno dei quali è formato da una forcina β (β – *hairpin*) di un monomero [19, 20, 21]. Questa è una caratteristica propria delle proteasi (retro)virali, mentre nelle proteasi degli eucarioti il sito attivo è coperto da un solo flap.

²Un nucleoside è una base azotata legata a una molecola di zucchero, necessita quindi dell'aggiunta di un gruppo fosfato (fosforilazione) per diventare un nucleotide. Avvenuto questo processo, gli inibitori nucleosidici agiscono sostituendosi alle basi azotate durante la trascrizione, in modo che il DNA provirale neoformato sia incompleto e quindi incapace di originare nuove particelle virali.

³Gli inibitori nucleotidici si differenziano da quelli nucleosidici perché sono già attivati chimicamente.

⁴Il primo termine indica l'amminoacido e il numero corrisponde alla posizione nel primo monomero. Un apice sul numero indica che l'amminoacido appartiene al secondo monomero.

Nella proteasi di HIV i flap sono costituiti dai residui 45 – 55 e consistono di tre regioni caratteristiche:

- la regione esposta al solvente costituita dai residui *M46*, *F53*, and *K55*;
- la regione idrofobica interna fatta di *I47*, *I54*, and *I56*;
- la regione ricca di glicina composta da *G48*, *G49*, *G51*, *G52*.

La regione ricca di glicina in combinazione con *I50* forma le punte dei flap, costituite dai residui 48 – 51 [22].

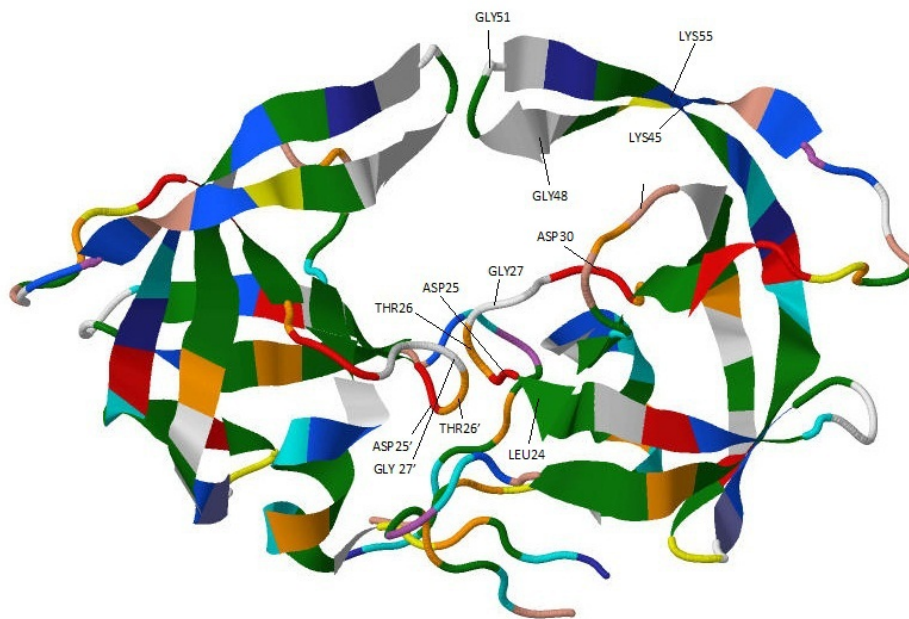


Figura 2.9: Strutture della proteasi di HIV-1, con evidenziati i diversi amminoacidi. Si può vedere la regione dei flap (residui 45 – 55), quella del sito attivo (residui 24 – 30) e la triade catalitica Asp25, Thr26, Gly27. (PDB: 1HHP)

Quando la proteasi si lega ad un substrato o ad un inibitore, è indotto un ampio moto dei flap e la loro apertura permette al substrato di entrare nel sito attivo [23]. Tuttavia, tramite simulazioni al computer è stato osservato che non c'è sempre bisogno dell'apertura dei flap per far entrare il substrato: piccole molecole (ad esempio inibitori) possono entrare nel sito attivo attraverso una cavità laterale [23, 24].

Il meccanismo catalitico delle proteasi aspartiche segue diversi passi. Prima si ha il legame della proteasi con il substrato per formare un complesso legato [24]. I flap chiudono l'ingresso al sito attivo instaurando la corretta geometria per permettere la proteolisi del substrato [25]. Successivamente i prodotti ottenuti dal taglio del substrato vengono rilasciati. È stato osservato in simulazioni di dinamica molecolare che il rilascio dei prodotti non richiede obbligatoriamente l'apertura dei flap perché possono scivolare fuori dai lati del sito di taglio [26].

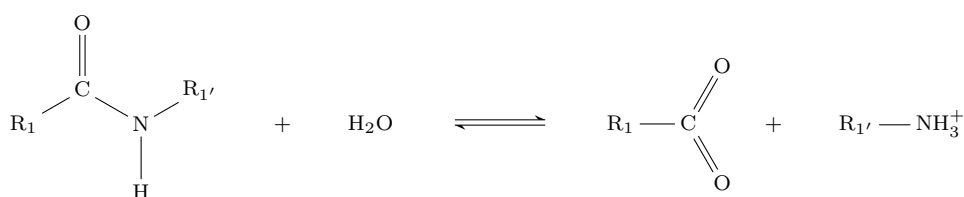
Inoltre è stato osservato che nella proteasi di HIV i flap possono esistere in differenti stati conformazionali, molte strutture adottano una conformazione semi-aperta nella quale le punte dei flap si allontanano dal sito di legame, si ha una conformazione chiusa ed esiste anche una conformazione totalmente aperta [22]. Da esperimenti NMR è stato suggerito che le conformazioni semi-aperte, chiuse e completamente aperte sono in equilibrio dinamico nella scala temporale dei microsecondi, mentre le punte dei flap presentano fluttuazioni nella scala dei tempi dei nanosecondi [27].

L'elevato tasso di mutazione proprio dei retrovirus influisce anche sui diversi siti della proteasi causando vari polimorfismi⁵. Delle 99 posizioni presenti in ciascun monomero, 37 sono note come invarianti (con frequenze di mutazione < 0,5%) e 17 sono siti di polimorfismi non dovuti alle terapie antivirali [29, 30]. Si ritiene invece che 45 posizioni in ciascun monomero siano coinvolte nella resistenza ai farmaci (vedi figura 2.10). Di queste 45 posizioni, è stato mostrato che 46 mutazioni in 25 siti diminuiscono significativamente in suscettibilità ad uno o più inibitori della proteasi mentre le altre sono mutazioni polimorfiche che si verificano più frequentemente quando sono associate alla terapia con inibitori. Le 46 mutazioni causanti la diminuzione significativa della suscettibilità agli inibitori sono L10FI⁶, V11L, K20T, L24FI, D30N, V32I, L33F, E35GN, K43T, M46IL, I47AV, G48MV, I50LV, F53L, I54ALMSTV, Q58E, G73CST, T74PS, L76V, V82AFLST, N83D, I84AV, N88DS, L89V, e L90M [31]. Inoltre, quasi il 60% di queste 25 posizioni cade al di fuori della regione del sito attivo [29].

Gli studi sulla specificità della proteasi sono importanti allo scopo di progettare suoi inibitori e per questo motivo si studia come predire l'idrolisi di un substrato da parte della proteasi. Il fatto che il virus muti anche al di fuori del sito attivo indica che il meccanismo di azione della proteasi ha carattere di non località e spinge a cercare strategie inibitorie di tipo allosterico, cioè che coinvolgono siti della proteina anche lontani dal sito attivo.

2.6.1 La reazione di idrolisi

La proteasi può tagliare un gran numero di proteine del substrato grazie ad una reazione di idrolisi:



Nonostante l'idrolisi dei legami peptidici sia favorita termicamente, è un processo a barriera di attivazione molto alta che avverrebbe in 1000 anni in assenza di catalizzatori, i quali riducono notevolmente i tempi di reazione portandoli all'ordine dei millisecondi [32].

⁵Per polimorfismo si intende la variabilità naturale esistente in una certa posizione del genoma con una frequenza nella popolazione maggiore dell'1%, altrimenti viene classificata come una semplice mutazione [28].

⁶la prima lettera (in questo caso L) indica l'amminoacido naturale, il numero (10) la posizione occupata nella proteina e le lettere seguenti (FI) gli amminoacidi in cui può occorrere la mutazione. La nomenclatura relativa ai vari amminoacidi è data dal codice a una lettera ripetuto nell'appendice A.

Tabella sulle frequenze di mutazione dei siti della proteasi in base al numero di PI ricevuti [30, 31]

Posizione	aa	Frequenza di mutazione al variare del numero di PI					mutazioni frequenti	caratteristica del sito
		0	1	2	3	≥4		
1	P	0.0	0.0	0.0	0.0	0.0		invariante
2	Q	0.2	0.2	0.6	0.0	0.0		invariante
3	I	0.4	0.0	2.2	0.0	0.0		invariante
4	T	0.1	0.0	0.6	0.4	0.5		invariante
5	L	0.0	0.5	0.6	0.4	0.0		invariante
6	W	0.1	0.3	0.6	0.9	0.5		invariante
7	Q	0.2	0.2	0.0	0.9	0.0		invariante
8	R	0.3	0.2	0.0	0.4	0.0		invariante
9	P	0.2	0.0	0.0	0.0	0.5		invariante
10	L	11.6	33.8	52.2	69.2	77.2	IVFR	resistente ai farmaci
11	V	0.2	0.6	1.1	3.0	3.7	I	resistente ai farmaci
12	T	11.6	8.3	7.2	6.0	10.6	SPAKEI	polimorfismo
13	I	12.5	21.0	23.9	15.8	15.9	V	mutazione indotta
14	K	11.0	12.2	10.0	9.0	9.5	R	polimorfismo
15	I	15.0	18.5	14.4	21.4	21.7	V	polimorfismo
16	G	2.6	2.8	2.2	2.1	3.7	EA	polimorfismo
17	G	0.3	0.3	0.6	0.4	0.5		invariante
18	Q	2.0	1.7	1.7	1.7	3.2	H	polimorfismo
19	L	8.6	9.7	13.3	9.4	13.2	IVTQ	polimorfismo
20	K	1.6	12.1	19.4	27.8	29.6	RIMT	resistente ai farmaci
21	E	0.2	0.0	0.0	0.0	0.0		invariante
22	A	0.0	0.6	0.0	1.7	1.6	V	mutazione indotta
23	L	0.1	0.6	2.2	3.8	1.6	I	mutazione indotta
24	L	0.1	5.2	5.6	10.7	9.5	I	resistente ai farmaci
25	D	0.0	0.2	0.0	0.9	0.0		invariante
26	T	0.0	0.0	0.0	0.0	0.0		invariante
27	G	0.0	0.0	0.0	0.0	0.0		invariante
28	A	0.0	0.0	0.0	0.9	0.0		invariante
29	D	0.3	0.0	0.0	0.4	0.0		invariante
30	D	0.0	13.8	12.2	3.4	3.2	N	resistente ai farmaci
31	T	0.1	0.2	0.0	0.0	0.0		invariante
32	V	0.0	4.2	3.9	3.8	6.3	I	resistente ai farmaci
33	L	2.9	4.7	4.4	7.3	14.8	FIV	resistente ai farmaci
34	E	0.5	0.9	2.8	3.0	4.8	QD	mutazione indotta
35	E	24.0	28.7	28.9	32.1	39.7	DG	resistente ai farmaci
36	M	13.3	27.8	30.6	33.8	44.4	IVL	mutazione indotta
37	N	34.6	34.5	31.1	35.0	35.4	DSTEHCA	polimorfismo
38	L	0.5	0.5	0.6	0.0	0.5		polimorfismo
39	P	3.2	3.5	1.1	1.3	2.6	SQ	polimorfismo
40	G	0.1	0.2	0.0	0.0	0.0		invariante
41	R	21.9	17.0	19.4	22.6	19.0	K	polimorfismo
42	W	0.0	0.0	0.0	0.0	0.5		invariante
43	K	2.1	2.8	7.8	7.3	7.9	TR	resistente ai farmaci
44	P	0.1	0.0	0.0	0.0	0.0		invariante
45	K	0.4	3.0	0.6	3.8	4.2	KR	mutazione indotta
46	M	0.2	22.4	38.9	42.7	51.3	ILV	resistente ai farmaci
47	I	0.3	2.2	1.1	1.7	3.2	V	resistente ai farmaci
48	G	0.1	3.5	5.6	12.8	10.1	V	resistente ai farmaci
49	G	0.1	0.0	0.0	0.0	0.0		invariante
50	I	0.1	1.4	0.6	0.9	1.4	V	resistente ai farmaci
51	G	0.1	0.3	0.6	0.9	0.0		invariante
52	G	0.1	0.2	0.0	0.0	0.0		invariante
53	F	0.4	1.4	5.0	11.1	15.3	LY	resistente ai farmaci
54	I	0.3	13.2	20.6	42.7	46.0	VLTM	resistente ai farmaci
55	K	0.7	2.2	4.4	3.4	7.4	R	mutazione indotta
56	V	0.0	0.0	0.0	0.0	0.0		invariante

continua nella pagina seguente

continuazione Tabella in figura 2.10

57	R	7.5	8.5	8.3	8.1	15.9	K	polimorfismo
58	Q	0.3	2.7	3.9	6.8	7.4	E	resistente ai farmaci
59	Y	0.3	0.2	0.0	0.0	0.0		invariante
60	D	3.7	6.4	7.2	13.7	13.8	E	mutazione indotta
61	Q	3.6	2.7	6.7	5.1	5.8	ENH	polimorfismo
62	I	17.2	27.8	32.8	36.3	44.4	V	mutazione indotta
63	L	68.0	82.6	90.6	91.5	96.8	PSAQTCHV	mutazione indotta
64	I	21.2	25.9	18.9	22.6	21.2	VLM	polimorfismo
65	E	2.0	2.0	2.2	1.3	1.1	D	polimorfismo
66	I	0.1	1.4	3.9	2.6	3.2	FV	mutazione indotta
67	C	1.9	2.4	3.3	1.7	5.3	SFY	polimorfismo
68	G	0.2	0.0	0.6	0.5	0.4		invariante
69	H	5.4	5.2	7.2	6.0	11.1	HKYR	polimorfismo
70	K	3.5	3.3	2.2	4.3	1.6	RQ	polimorfismo
71	A	6.8	34.4	50.0	57.3	69.3	VTI	mutazione indotta
72	I	8.3	15.2	13.9	21.8	16.4	VTMLE	mutazione indotta
73	G	0.0	6.1	14.4	27.4	34.9	STC	resistente ai farmaci
74	T	0.2	6.6	9.4	7.7	8.5	SPA	resistente ai farmaci
75	V	0.1	1.3	1.7	2.1	0.0	I	mutazione indotta
76	L	0.1	1.3	2.2	1.3	0.5	V	resistente ai farmaci
77	V	23.4	28.6	36.7	34.6	34.9	I	mutazione indotta
78	G	0.0	0.0	0.0	0.0	0.0		invariante
79	P	0.0	0.5	1.7	1.7	4.2	A	mutazione indotta
80	T	0.0	0.0	0.0	0.0	0.0		invariante
81	P	0.0	0.2	0.0	0.0	0.0		invariante
82	V	1.2	25.0	33.3	46.2	42.3	ATFI	resistente ai farmaci
83	N	0.0	0.5	1.1	0.9	1.6	D	resistente ai farmaci
84	I	0.0	5.0	16.1	24.8	36.5	V	resistente ai farmaci
85	I	0.2	2.8	1.7	4.7	5.7	V	mutazione indotta
86	G	0.0	0.2	0.0	0.0	0.0		invariante
87	R	0.6	0.0	0.0	1.3	0.0		invariante
88	N	0.2	9.1	8.9	6.8	7.4	DS	resistente ai farmaci
89	L	0.7	1.4	5.0	7.3	4.2	MVI	resistente ai farmaci
90	L	0.1	22.3	40.6	55.6	66.7	M	resistente ai farmaci
91	T	0.6	0.2	0.0	0.4	0.0		invariante
92	Q	0.9	2.4	3.9	3.8	3.2	KR	mutazione indotta
93	I	24.1	34.5	36.7	42.3	48.7	LM	mutazione indotta
94	G	0.0	0.0	0.0	0.0	0.0		invariante
95	C	0.3	0.6	2.2	3.8	3.2	F	mutazione indotta
96	T	0.0	0.0	0.0	0.0	0.0		invariante
97	L	0.1	0.0	0.0	0.0	0.0		invariante
98	N	0.3	0.2	0.0	0.0	0.0		invariante
99	F	0.0	0.0	0.6	0.0	0.0		invariante

Figura 2.10: La prima colonna si riferisce alla posizione dell'amminoacido in un monomero della proteasi di HIV-1 sottotipo B, nella seconda colonna è riportato l'amminoacido relativo alla posizione del sito della proteasi nativa utilizzando la nomenclatura ad una lettera (vedi appendice A), a seguire la frequenza di mutazioni (in percentuale) al variare del numero dei tipi di inibitori della proteasi utilizzati dal paziente, nell'ottava colonna sono evidenziate le mutazioni più frequenti avvenute nello specifico sito e nell'ultima è riportato se il sito è invariante oppure se la mutazione è dovuta a polimorfismo (riga evidenziata di giallo), se si verifica più frequentemente quando è associata alla terapia con inibitori (riga evidenziata di verde) o se nel sito è indotta almeno una mutazione causata dalla resistenza ai farmaci (riga evidenziata di azzurro).

2.6.2 Sequenze tagliate dalla proteasi dell'HIV

La proteasi è in grado di idrolizzare un gran numero di octapeptidi che si trovano in diverse proteine come si può vedere dalla tabella in figura 2.11.

		domini tagliati	sequenza di taglio
PROTEINE VIRALI	GAG	MA-CA	SQNYPIVQ
		CA(CTD)-CA(NTD)	AEAMSQVT
		CA-SP1	ARVLAEAM
		SP1-NC	TAIMMQKG
			SAIMMQRG
			ATIMMQRG
		NC-SP2	RQANFLGK
		SP2-p6	PGNFLQSR
	p6	FRSGVETT	
	POL	NC/TFP	RQANFLRE
		TFP/p6pol	NLAFQQGE
		TFP/p6pol	DLAFLOKKG
		p6pol/PR	SFSFPQIT
		p6pol/PR	SFNFPQVT
		PR/RTp51	TLNFPISP
		PR/RTp66	AETFYVDG
		RTp66/IN	RKVLFLDG
		Nef	DCAWLEAQ
Nef		ACAWLEAQ	
ALTRE PROTEINE	CALMODULINA	GQVNYEEF	
		YEEFVQMM	
		REAFRVFD	
	ACTINA	TQIMFETF	
		PTLLTEAP	
		SFIGMESA	
	INSULINA	ETTALVCD	
		GSHLVEAL	
	VIMENTINA	HLVEALYL	
		SLNLRETN	
		DAINTEFK	
		SRSLYASS	
GGVYATRS			

Figura 2.11: Nella tabella sono riportate alcune sequenze tagliate dalla proteasi di HIV-1 e il loro rispettivo dominio, mostrando che la proteasi può tagliare anche delle proteine che non hanno alcuna correlazione con il virus HIV [21, 33].

Risulta utile cercare di predire le sequenze tagliate ed eventualmente la probabilità di taglio anche allo scopo di chiarire i meccanismi di riconoscimento del sito di taglio e di progettare inibitori peptidici o peptidomimetici⁷ (nella Tabella A.1 sono illustrati alcuni lavori effettuati al fine di ottenere questa predizione ed i principali risultati ottenuti in questi lavori sono riportati in appendice B). Trovare un metodo che permetta a priori di distinguere quali peptidi possano o meno essere tagliati dalla proteasi è importante a causa dell'alto numero di possibili octapeptidi, 2.56×10^{10} , che possono essere creati dai 20 amminoacidi a disposizione.

Uno studio significativo è quello di Chou [21]. In questo studio viene fornito un algoritmo per la predizione della probabilità di taglio di una data sequenza octapeptidica. Gli otto residui del peptide sono sequenzialmente indicati da $R_4, R_3, R_2, R_1, R_{1'}, R_{2'}, R_{3'},$ ed $R_{4'}$ ed il legame viene tagliato tra R_1 ed $R_{1'}$ (vedi figura 2.12). Tuttavia il sito può contenere a volte

⁷ composti con caratteristiche strutturali analoghe a quelle dei peptidi naturali ma con migliori proprietà farmacologiche [34].

anche un eptapeptide o un nonapeptide [21]. L'apparato catalitico è costituito da gruppi carbossilici derivati da due residui aspartici (provenienti da Asp25 e Asp25') situati nelle metà terminali delle molecole N e C dell'enzima e successivamente vedremo come avviene il fenomeno dell'idrolisi all'interno del sito attivo.

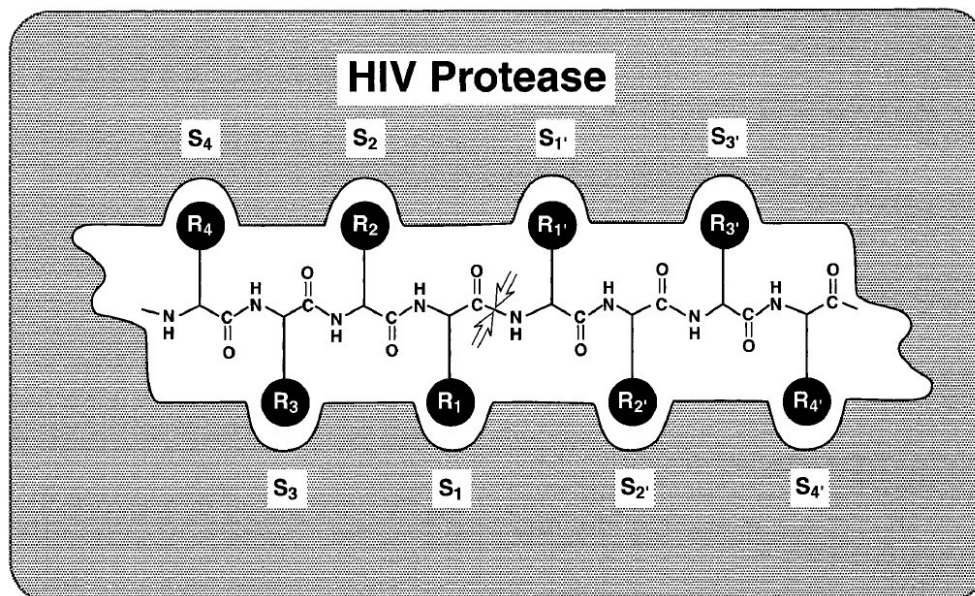


Figura 2.12: Rappresentazione schematica del legame tra il substrato e la proteasi di HIV. Il sito attivo dell'enzima è composto da otto sottositi $S_4, S_3, S_2, S_1, S_1', S_2', S_3', S_4'$ ed il substrato è rappresentato da $R_4, R_3, R_2, R_1, R_1', R_2', R_3', R_4'$. La scissione del legame avviene tra R_1 e R_1' [21].

2.6.3 L'ipotesi dell'involuppo del substrato

Un'ipotesi complementare al riconoscimento sequenziale del substrato da parte di PR è quella del riconoscimento strutturale/chimico-fisico. Da questa ipotesi deriva una tecnica alternativa per la progettazione di inibitori basata sulla costruzione dell'involuppo del substrato (*substrate envelope*) [42]. Studi cristallografici hanno rivelato che una caratteristica strutturale comune dei substrati naturali è il volume che essi adottano all'interno del sito di legame dell'enzima. Questo volume è definito come involuppo del substrato, e si ipotizza che sia l'elemento chiave per il riconoscimento del substrato piuttosto che una specifica sequenza di amminoacidi. La sovrapposizione dell'inibitore e del substrato ha mostrato che gli inibitori si protraggono oltre l'involuppo del substrato e fanno contatto con alcuni altri residui della proteasi. Questi residui corrispondono ad alcuni dei siti di mutazione di resistenza ai farmaci nella proteasi [42], precedentemente mostrati in figura 2.10. L'ipotesi dell'involuppo del substrato afferma che gli inibitori della proteasi che si adattano completamente all'interno dell'involuppo del substrato probabilmente sono meno sensibili alle mutazioni di resistenza ai farmaci rispetto a quelli che superano l'involuppo e forniscono il modo di ottenere mutazioni che abbassano l'affinità con l'inibitore [43, 44]. In seguito a questa ipotesi, l'involuppo del substrato è stato usato con successo come un vincolo per guidare la progettazione di nuove molecole che sono efficaci inibitori della proteasi. Infatti esplorando l'involuppo del substrato

Autore	anno	referenza	descrizione
Chou	1996	[21]	siti di taglio predetti dalla struttura primaria della proteina includendo sia gli effetti tra i sottositi con posizioni adiacenti che quelli con posizioni alternate classificatore non lineare
Rögnvaldsson e You	2004	[35]	per predire i siti di taglio è più opportuno utilizzare i classificatori lineari rispetto a quelli non lineari
Liwen You	2006	[36]	predizione siti di taglio nelle proteine native non virali combinando le informazioni ottenute da brevi oligopeptidi, strutture secondarie e dall'accessibilità del solvente
Bing Niu, Lin Lu et al	2008	[37]	siti di taglio predetti dalla struttura secondaria e da altre caratteristiche
Rögnvaldsson et al	2009	[38]	si guarda la struttura primaria del substrato che per essere tagliato deve alcune avere determinati amminoacidi in determinate posizioni
Hasan Ogul	2009	[39]	il processo di legame della proteasi oltre che dalle interazioni tra la proteasi e il substrato include le interazioni dei residui della sequenza del sito di taglio
Hyeoncheol Kim et al	2010	[40]	selezione del substrato con caratteristiche rilevanti prese da un dominio non lineare
Nanni e Lumini	2010	[41]	rappresentazione matriciale del substrato costruita tramite le proprietà chimicofisiche degli amminoacidi

Tabella 2.1: Lista di lavori riguardanti la predizione delle sequenze di amminoacidi tagliate dalla proteasi di HIV-1. Nell'appendice B sono riportati i principali risultati ottenuti in questi lavori.

per progettare un inibitore della proteasi di HIV-1 si imita la forma del substrato [44] (vedi figura 2.14). Secondo l'ipotesi dell'inviluppo del substrato, i substrati che sporgono oltre l'inviluppo e prendono contatto con i siti di mutazione di resistenza ai farmaci della proteasi co-evolvono con le varianti della proteasi [42].

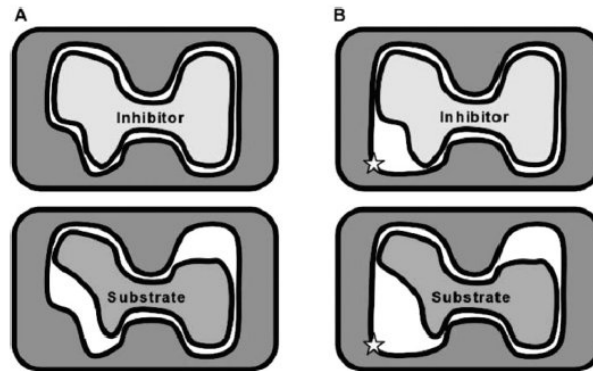


Figura 2.13: Illustrazione dell'ipotesi dell'inviluppo del substrato. Nel bersaglio dei farmaci di tipo naturale, l'inibitore tradizionale (A, in alto) occupa più del sito di legame e fa più contatti di un substrato (A, in basso). In B, il bersaglio dei farmaci è mutato per espandere il sito attivo in una regione che solo contatta l'inibitore (stella). L'inibitore (B, in alto) perde contatti e di conseguenza l'affinità di legame, mentre il substrato (B, in basso) perde affinità in maniera trascurabile [44].

Che la proteasi riconosca una forma asimmetrica adottata dai peptidi del substrato è mostrato anche dal fatto che i siti di taglio della proteasi di HIV-1 nelle poliproteine Gag e Gag-Pol sono non omologhi e asimmetrici, sia nella carica che nelle dimensioni [29].

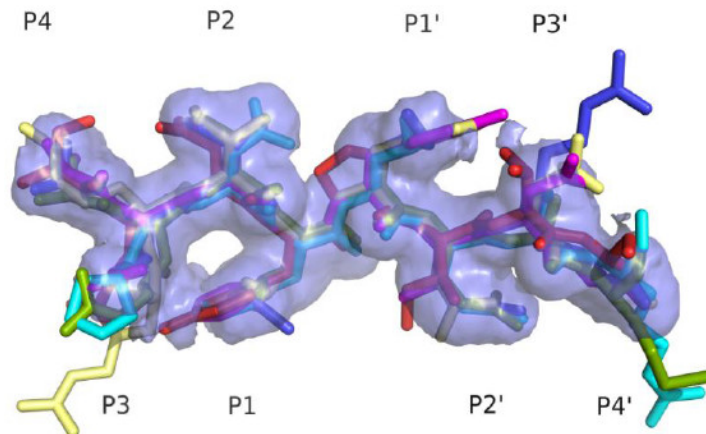


Figura 2.14: Inviluppo del substrato della proteasi di HIV generato dalla sovrapposizione dei volumi dati dai legami van der Waals dei seguenti substrati. Rosso: MA-CA, verde: CA-p2, blu: p2-NC, ciano: p1-p6, magenta: trascrittasi inversa-ribonucleasi, giallo: ribonucleasi-integrasi [29].

2.6.4 Conclusioni

Dalla lettura dei vari lavori visti precedentemente, si evince che il riconoscimento del substrato da parte della proteasi di HIV-1 appare più basato sulla struttura del substrato che sulla sequenza, anche se questa è un'ipotesi ancora da verificare.

Sembra importante non solo la struttura della parte del substrato che interagisce con il sito di taglio, ma anche il ripiegamento circostante del substrato. Per questo motivo in questo lavoro di tesi si mira ad uno studio dinamico dell'interazione tra PR e substrato intero e ad analizzare in particolare le interazioni locali che questo altera non solo nel sito attivo, ma anche nelle regioni circostanti, soprattutto quelle ad alta variabilità.

La sequenza del substrato appare invece più importante per la probabilità che avvenga la reazione idrolisi da parte di PR.

Capitolo 3

Dinamica molecolare classica con modelli a bassa risoluzione

3.1 Dinamica molecolare classica

Le simulazioni al computer tramite dinamica molecolare classica (MD) sono da tempo diventate uno strumento largamente usato nello studio dei sistemi macromolecolari biologici e in generale nella biofisica. Nell'ambito della MD classica dopo aver costruito un modello teorico del sistema (un'hamiltoniana), il sistema viene evoluto al computer, cioè si risolvono numericamente, a partire da condizioni iniziali date (configurazione iniziale e velocità), equazioni basate su quella di Newton, le quali possono ben approssimare grandi sistemi [45]:

$$m_i \ddot{x}_i = - \frac{\partial U(x_1, \dots, x_n)}{\partial x_i} \quad (3.1)$$

dove $\partial U / \partial x_i$ è la forza esercitata sulla particella avente massa m_i ed accelerazione \ddot{x}_i . La forza è il gradiente di un'energia potenziale, identificata in questo caso come una superficie di energia potenziale (PES), dalla quale vengono ricavate le varie accelerazioni atomiche e quindi le traiettorie atomiche.

3.1.1 Algoritmi di integrazione

Per risolvere l'equazione di Newton si usa una procedura di integrazione numerica. Date le posizioni e le velocità in uno stato iniziale t , quelle al tempo $t + \Delta t$ si possono ottenere come una loro funzione tramite uno sviluppo in serie di Taylor:

$$r(t + \Delta t) = r(t) + \dot{r}(t)\Delta t + \frac{1}{2} \frac{F(t)}{m} \Delta t^2 + O(t^3) \quad (3.2)$$

con un errore numerico del terzo ordine in Δt . Per migliorare l'errore numerico, si usa in realtà un algoritmo leggermente diverso ma sempre basato sull'espansione di Taylor, chiamato algoritmo di Verlet:

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \frac{F(t)}{m} \Delta t^2 + O(t^4) \quad (3.3)$$

che permette di avere un errore del quarto ordine in Δt . Per raggiungere grandi scale temporali nella simulazione, si cerca di avere il timestep Δt il più grande possibile, in quanto il costo computazionale che è dato dal numero totale di passi temporali. Tuttavia per limitare l'errore numerico introdotto a ciascun passo dell'integrazione, bisogna scegliere Δt molto piccoli, facendo risultare la scelta del timestep critica nel determinare la qualità della simulazione [45, 46].

Per quanto riguarda la propagazione delle velocità, andando a sottrarre al termine $r(t + \Delta t)$ il termine $r(t - \Delta t)$ otteniamo:

$$v_n = \frac{r(t + \Delta t) - r(t - \Delta t)}{2\Delta t} + O(t^2) \quad (3.4)$$

Per migliorare il calcolo delle velocità, dall'algoritmo di Verlet sono stati sviluppati altri due metodi di integrazione: l'integratore Verlet Leapfrog e l'integratore Velocity Verlet.

Nel Verlet Leapfrog il primo passo è quello di calcolare la velocità al tempo $t + (1/2)\Delta t$ tramite l'integrazione della forza:

$$v(t + \frac{1}{2}\Delta t) = v(t - \frac{1}{2}\Delta t) + \frac{F(t)}{m}\Delta t \quad (3.5)$$

e le posizioni sono calcolate usando le nuove velocità:

$$r(t + \Delta t) = r(t) + v(t + \frac{1}{2}\Delta t)\Delta t \quad (3.6)$$

Poiché le simulazioni di dinamica molecolare richiedono proprietà che abbiano posizioni e velocità allo stesso tempo, viene calcolata la velocità al tempo t dalla media delle velocità al tempo $(t + (1/2)\Delta t)$ e $(t - (1/2)\Delta t)$:

$$v(t) = \frac{1}{2} \left[v(t - \frac{1}{2}\Delta t) + v(t + \frac{1}{2}\Delta t) \right] \quad (3.7)$$

Un miglioramento nel calcolo delle velocità è ottenuto da un'altra variante dell'integratore di Verlet, l'algoritmo velocity Verlet. Questo algoritmo assume che le posizioni, le velocità e le forze siano conosciute a ogni passo temporale. La velocità è calcolata usando l'equazione

$$v(t + \frac{1}{2}\Delta t) = v(t) + \frac{1}{2} \frac{F(t)}{m} \Delta t \quad (3.8)$$

quindi si ottiene la posizione

$$r(t + \Delta t) = r(t) + v(t + \frac{1}{2}\Delta t)\Delta t \quad (3.9)$$

Usando la nuova posizione al tempo $(t + \Delta t)$, si ottiene la forza al tempo $(t + \Delta t)$, grazie alla quale si può adesso calcolare la velocità

$$v(t + \Delta t) = v(t + \frac{1}{2}\Delta t) + \frac{1}{2} \frac{F(t + \Delta t)}{m} \Delta t \quad (3.10)$$

Si è così ottenuto il totale sincronismo tra posizioni, velocità e forze [45, 47].

3.1.2 Termostati

La dinamica newtoniana descrive il comportamento di un sistema ad energia totale costante di N particelle in un volume V , l'ensemble microcanonico (NVE). A volte però si è interessati a sistemi che abbiano una data temperatura, come l'ensemble canonico (NVT), oppure a un ensemble isotermico-isobaro (NPT) [45]. Le simulazioni NVT necessitano di un termostato e il metodo più semplice per mantenere costante la temperatura consiste nel riscaldare ad ogni passo tutte le velocità di un fattore $\lambda = (T_0/T)^{1/2}$, dove T è la temperatura istantanea e T_0 è la temperatura di riferimento. Questo metodo però ha il limite di introdurre variazioni troppo brusche delle velocità e può portare ad una dinamica troppo distorta.

Un metodo più raffinato si basa sul termostato di Berendsen. In questo caso il sistema viene accoppiato con un bagno termico esterno avente una temperatura T_0 , con una dinamica del primo ordine con tempo di rilassamento τ . Il cambio di temperatura tra due successivi passi temporali è:

$$\Delta T = \frac{\Delta t}{\tau}(T_0 - T(t)) \quad (3.11)$$

Le velocità sono così riscaldate del fattore λ nuovamente definito come:

$$\lambda = \left[1 + \frac{\Delta t}{\tau} \left(\frac{T_0}{T(t - \Delta t/2)} - 1 \right) \right]^{1/2} \quad (3.12)$$

Per il valore $\tau = \infty$ si riottiene il campionamento microcanonico, se invece si pone $\Delta t = \tau$ si ritrova il semplice metodo di riscalaggio visto precedentemente. Tuttavia l'ensemble generato utilizzando il termostato di Berendsen non è un ensemble canonico, per ottenere il quale bisogna utilizzare il termostato di Nose-Hoover.

Il termostato di Nose-Hoover stabilisce un accoppiamento lagrangiano al bagno termico di un nuovo grado di libertà che descrive le fluttuazioni del termostato, la variabile s che è un fattore di riscalaggio delle velocità associata a una massa $Q > 0$. La lagrangiana per il nuovo sistema è:

$$L = \sum_i \frac{m_i}{2} s^2 v_i^2 - E(r_i) + \frac{1}{2} Q \frac{\dot{s}^2}{s^2} - gkT_0 \ln(s) \quad (3.13)$$

I primi due termini rappresentano l'energia cinetica e l'energia potenziale del sistema reale, il terzo l'energia cinetica di s e il quarto il potenziale avente il parametro di accoppiamento col bagno termico g . Le equazioni lagrangiane possono essere riscritte introducendo la variabile $\gamma = \dot{s}/s$:

$$m_i \ddot{r}_i = F_i - m_i \gamma \dot{r}_i \quad (3.14)$$

$$\ddot{\gamma} = -\frac{Nk_B}{Q} T(t) \left(\frac{g}{N} \frac{T_0}{T(t)} - 1 \right) = -\frac{1}{\tau_{NH}^2} \left(\frac{T_0}{T(t)} - 1 \right) \quad (3.15)$$

dove il tempo di rilassamento è:

$$\tau_{NH}^2 = \frac{Q}{Nk_B T(t)} \quad (3.16)$$

Il significato delle equazioni 3.14 e 3.15 è quello di applicare al sistema una nuova equazione del moto con un termine di tipo viscoso ma con un coefficiente di attrito γ che varia con il tempo e che è regolato dal discostamento della temperatura da quella di riferimento. Grazie al termostato di Nosé-Hoover si ha un campionamento più accurato dell'ensemble canonico [45, 46, 48].

3.1.3 Dinamica di Langevin e browniana

Abbiamo così visto dinamiche di tipo deterministico, dove la traiettoria percorsa è completamente determinata dalle condizioni iniziali del sistema. Nel caso in cui ci siano gradi di libertà nascosti, dinamiche di tipo stocastico sono più realistiche. In queste condizioni si può utilizzare la dinamica di Langevin:

$$m_i \ddot{r}_i = F_i - m_i \gamma_i \dot{r}_i + F_i^s \quad (3.17)$$

Il secondo termine della parte destra dell'equazione rappresenta il termine viscoso che sottrae energia al sistema, proporzionale alle velocità \dot{r}_i e ai coefficienti di attrito γ . Il terzo termine è una forza random che invece aggiunge energia al sistema. Affinché il valore medio dell'energia rimanga costante, tra la funzione di correlazione temporale della forza stocastica e il coefficiente di attrito deve valere la seguente relazione:

$$\langle F_i^s(0) F_i^s(t) \rangle = 2m_i \gamma_i k_B T \delta_{ij} \quad (3.18)$$

La delta di Dirac presente al secondo termine indica che gli urti sono temporalmente scorrelati tra loro.

La dinamica di Langevin può essere vista come un termostato. Per grandi valori di γ si può trascurare il termine di secondo ordine dell'equazione di Langevin che si trasforma nell'equazione della dinamica Browniana:

$$dr_i = \frac{D_i}{k_B T} F_i dt + f_i^s \quad (3.19)$$

dove D è il coefficiente di diffusione che è legato al coefficiente d'attrito

$$\gamma = \frac{k_B T}{mD} \quad (3.20)$$

Le equazioni di dinamica stocastica e in particolare l'equazione di Langevin sono utili per reintrodurre in maniera implicita l'effetto dinamico di gradi di libertà non trattati esplicitamente. Un caso tipico riguarda le trattazioni con solvente implicito [45, 46, 48].

3.2 Campi di forze atomistici

Le simulazioni di MD a livello atomico comprendono i metodi di meccanica quantistica (QM), in cui gli elettroni sono trattati esplicitamente e ovviamente quantisticamente in approssimazione di Born Hoppenheimer, mentre i nuclei sono generalmente oggetti classici e trattati con metodi di meccanica molecolare (MM), in cui si eliminano i gradi di libertà

elettronici. Questo è necessario in sistemi di dimensioni grandi come le proteine, dal momento che con le simulazioni a livello QM è impossibile affrontare lo studio di sistemi con un numero di atomi maggiore di 100-1000, almeno su scale di tempi interessanti. È quindi necessario snellire il sistema. La base dei vari metodi riguarda sempre la PES:

- a livello QM la PES viene calcolata esattamente dall'equazione di Schrodinger;
- a livello MM la PES viene approssimata con potenziali empirici (Campo di Forze)

In un campo di forze empirico (chiamato anche *force field* o FF) la PES viene approssimata con dei termini analitici, dipendenti dal tipo di atomo coinvolto nell'interazione, in quanto i vari elementi hanno diversi tipi di atomi, e dall'ambiente chimico circostante. A questi termini analitici si cerca di dare un significato fisico tramite le forze presenti all'interno delle proteine, in quanto tengono conto sia delle interazioni di tipo legato che di quelle di tipo non legato:

$$\begin{aligned}
 E = & \sum_{\text{legami}} k_b (d - d_0)^2 + \sum_{\text{angoli}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{diedri}} k_\phi [1 + \cos(n\phi + \delta)] \\
 & + \sum_{\substack{\text{coppie non} \\ \text{legate } AB}} \left\{ \varepsilon_{AB} \left[\left(\frac{\sigma_{AB}}{r_{AB}} \right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}} \right)^6 \right] + \frac{1}{4\pi\varepsilon_0} \frac{q_A q_B}{r_{AB}} \right\} \quad (3.21)
 \end{aligned}$$

I primi tre termini nella parte destra dell'equazione 3.21 rappresentano le interazioni tra atomi chimicamente legati: il primo termine riguarda i possibili legami chimici del sistema, il secondo tutti i possibili 'angoli di legame' formati da triplette di atomi legati, il quarto riguarda gli angoli diedri formati da quattro atomi legati. I termini successivi invece rappresentano interazioni tra atomi non direttamente legati tra loro: il quarto termine descrive le interazioni di Van Der Waals e il quinto il potenziale elettrostatico.

3.3 Campi di forze a bassa risoluzione

Tuttavia anche i metodi di MD classica basati sui campi di forza empirici presentano dei limiti. Risulta difficile studiare aggregati macromolecolari, le cui dimensioni sono dell'ordine di $10nm - 1\mu m$, e c'è un limite sulle scale temporali dell'ordine dei $100ns$ fino alle decine dei μs per le piccole proteine. Le scale spatio-temporali possono essere estese di qualche ordine di grandezza su sistemi massicciamente paralleli. Tuttavia la necessità di scale temporali più lunghe e il dover rappresentare oggetti biologici di maggiori dimensioni ha portato alla creazione dei modelli coarse-grained (CG), cioè a bassa risoluzione.

Nei modelli CG alcuni gradi di libertà vengono inglobati in singoli centri interattivi chiamati *bead*, semplificando la superficie di energia libera e riducendo la dimensionalità e la complessità (e quindi il costo computazionale) del modello. A seconda del numero e collocazione dei centri interattivi, esistono vari tipi di modelli coarse-grained. Aumentando il livello di coarse-graining, si ha un maggior guadagno nel tempo di calcolo ma ci sono anche

degli svantaggi: la minore risoluzione produce generalmente una minore accuratezza, una minore trasferibilità del campo di forze ed una maggiore difficoltà nella parametrizzazione.

Nel seguito sarà descritto un particolare livello di risoluzione per proteine, detto one bead (OB), cioè un singolo centro interattivo per amminoacido, che quindi descrive la proteina come una catena lineare. Nei modelli OB, i campi di forza vengono scritti come somma di termini legati e non legati come avviene per i campi di forza MM:

$$U = U_b + U_\theta + U_\phi + U_{nb} \quad (3.22)$$

Nell'equazione 3.22, i primi tre termini nella parte destra rappresentano i termini legati, rispettivamente il legame bead-bead, il potenziale dell'angolo tra tre bead successivi e il potenziale del diedro tra quattro bead successivi. Il termine U_{nb} rappresenta i termini non legati (*non bonded*), e contiene l'effetto dei legami a idrogeno, la parte idrofobica e sterica e la parte elettrostatica.

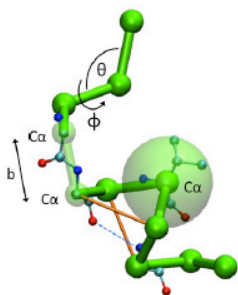


Figura 3.1: Modello a un bead posto sugli atomi C_α rappresentati da sferette di colore verde. La sfera più larga rappresenta l'interazione non-bonded. In arancione sono rappresentati i legami a idrogeno [49].

Un'ulteriore scelta da compiere sui modelli OB riguarda il posizionamento dei bead [49]. Le varie possibilità comprendono il centro di massa del bead, il centro geometrico del bead oppure la posizione specifica dei vari atomi inclusi nel bead. Nell'ultimo caso, le scelte più utilizzate ricadono sul C_β , ovvero il primo atomo di carbonio della catena laterale, oppure come è stato fatto anche in questo lavoro di tesi, il C_α (figura 3.1).

La scelta di prendere il C_α come bead, chiamato modello minimalista, è data dal fatto che comporta dei vantaggi soprattutto nella descrizione delle strutture secondarie e nella semplificazione dei potenziali dei termini legati. Il gruppo C_α -NH-CO- C_α è disposto su un piano e la distanza l , definita come la distanza C_α - C_α , è fissata ed indipendente dalla struttura secondaria e da quella terziaria. Invece l'angolo θ e il diedro ϕ descrivono la strutture secondarie e sono in corrispondenza biunivoca con essa. Così facendo l , θ e ϕ permettono un'accurata descrizione della struttura secondaria. Tuttavia, la scelta di prendere il C_α come bead, comporta una complicazione nei termini non legati presenti nel campo di forze, causando un'anisotropia nelle interazioni

che descrivono la parte idrofobica e il volume escluso, e le interazioni chimiche ed elettrostatiche. Diverse strategie sono state adottate nel tempo per risolvere questi problemi, come quelle dei modelli a rete e $G\ddot{o}$.

3.3.1 Modelli a rete elastica

Nei modelli a rete elastica (ENM) [50, 51], tutte le coppie di atomi ad una distanza minore di una distanza di taglio R_{cut} sono considerati legati. In questo modo spariscono i termini del potenziale relativi agli angoli e ai diedri e rimangono U_b e U_{nb} , e si ottiene un potenziale totale:

$$U = \sum_{i,j} u_b(|R_{ij} - R_{ij}^0|)(-\Gamma_{ij})$$

dove

$$u_b(|R_{ij} - R_{ij}^0|) = \frac{1}{2}k \sum_{i,j} (R_{ij} - R_{ij}^0)^2$$

è un potenziale armonico che dipende da una costante elastica k e dalla distanza di una coppia di atomi R_{ij} meno la loro distanza di equilibrio R_{ij}^0 ricavata da una struttura sperimentale, di solito cristallografica. Γ è conosciuta come matrice di Kirchhoff o di connettività [51] e indica che interagiscono tra di loro solo gli atomi all'interno di una sfera:

$$\Gamma_{ij} = \begin{cases} -1 & \text{se } i \neq j \text{ e } R_{ij}^0 \leq R_{cut}, \\ 0 & \text{se } i \neq j \text{ e } R_{ij}^0 > R_{cut}, \\ -\sum_{k \neq i} \Gamma_{ik} & \text{se } i = j. \end{cases}$$

I parametri che possono essere aggiustati sono R_{cut} e k , che viene fittato sulle fluttuazioni quadratiche medie sperimentali per ottenere delle fluttuazioni simili date dal modello teorico. Assumendo ulteriormente che le fluttuazioni abbiano una distribuzione gaussiana ed isotropa, modello a rete gaussiana, si può esprimere la correlazione delle fluttuazioni dell' i -esimo e del j -esimo atomo C_α (matrice di covarianza) come:

$$\langle \Delta R_i \Delta R_j \rangle = \frac{3k_B T}{k} \Gamma_{ij}^{-1}$$

con $\Delta R_i = R_i - R_i^0$, k_B è la costante di Boltzmann e T la temperatura assoluta. Diagonalizzando la matrice covarianza si ottengono gli autovalori che corrispondono ai modi del sistema, i cui principali sono quelli a frequenza più bassa che permettono di conoscere la natura fisiologica del sistema analizzato.

Reti bistabili

Lo svantaggio dei modelli a rete è dato dal non poter descrivere delle strutture e le dinamiche lontane dalla configurazione di riferimento, poiché si basano su una singola struttura di partenza. Così per sopperire a questa mancanza è stata considerata un'estensione a modelli a rete bistabile. In questo caso, è possibile partire da due strutture conosciute a priori, che permettono la creazione di potenziali a doppia buca con i minimi nelle due configurazioni di partenza.

Uno dei metodi sviluppati è il modello a rete elastica mista (MENM) [52], che permette di studiare le transizioni conformazionali a larga scala delle proteine, avendo a disposizione due o più strutture di partenza. Considerando il caso che sarà di interesse in seguito, cioè con due strutture note aventi l'insieme di coordinate rappresentato da $\{x_1\}$ e $\{x_2\}$ rispettivamente per la prima e per la seconda struttura, e indicando con $\{x\}$ l'insieme delle coordinate del sistema, abbiamo che l'energia totale del sistema è:

$$U(x) = -\frac{1}{k_B T_m} \ln \left[e^{-\frac{1}{k_B T_m} (U_1(x-x_1) + \epsilon_1)} + e^{-\frac{1}{k_B T_m} (U_2(x-x_2) + \epsilon_2)} \right]$$

con la temperatura di mixing T_m che determina l'altezza della barriera tra i due stati in relazione inversamente proporzionale. $U_1(x_1)$ e $U_2(x_2)$ sono le energie potenziali in appross-

simulazione ENM dei due stati ed ϵ_1 e ϵ_2 regolano le loro energie relative, che sono un input del modello.

Un metodo simile a quello appena visto è il modello a rete a doppia buca (DWNM) [53] dove si ottiene un risultato analogo al precedente anche se l'energia totale è scritta formalmente in modo differente:

$$U = \frac{U_1 + U_2 - \Delta}{2} - \frac{1}{2} \sqrt{(U_1 - U_2 + \Delta)^2 + \epsilon^2}$$

dove Δ è l'altezza relativa tra le due buche di potenziale ed ϵ è un termine indicante la curvatura della cima della barriera separante le due buche.

Grazie a questi due metodi è possibile passare da una conformazione ad un'altra, permettendo una rappresentazione più realistica del sistema analizzato poiché si può avere una maggiore varietà strutturale rispetto ai modelli visti precedentemente.

3.3.2 Modelli G \bar{o}

I modelli G \bar{o} [54] sono stati creati per studiare il folding delle proteine, con i parametri di equilibrio che sono ricavati dalla struttura nativa. Questo comporta un bias totale verso una singola struttura, come già avveniva nei precedenti modelli a rete. Il campo di forze torna ad essere simile a quello atomistico, con i termini legati dipendenti dalle distanze tra i bead, dagli angoli e dai diedri:

$$U_b = \sum_{i=1}^n \frac{1}{2} k_{b_i} (d - d_{0_i})$$

$$U_\theta = \sum_{i=1}^n \frac{1}{2} k_{\theta_i} (\theta - \theta_{0_i})$$

$$U_\phi = \begin{cases} \sum_{i=1}^n \frac{1}{2} k_{\phi_i} (\phi_i - \phi_{0_i}) \\ \text{oppure} \\ \sum_{i=1}^n k_{\phi_i} (1 + \cos n(\phi_i - \phi_{0_i})) \end{cases}$$

La parte di potenziale non bonded è formata da due tipi di potenziali, una parte locale ed una non locale divise da una distanza limite come per i modelli a rete. Per descrivere la parte locale si usa o un potenziale di Lennard Jones

$$\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right]$$

oppure un potenziale di tipo Morse

$$\epsilon \left[\left(e^{(-\alpha(d-d_0))} - 1 \right)^2 - 1 \right]$$

Il termine non locale invece rappresenta un potenziale repulsivo:

$$\epsilon\left(\frac{d_0}{d}\right)^{12}$$

Per fare in modo che la temperatura di folding del modello coincida con la temperatura di folding nota del sistema reale, viene fittato il parametro ϵ finché le due temperature non coincidono. L'unico minimo dell'energia libera in questo modello è dato dalla struttura nativa, il che causa la frustrazione minima. Questo è un vantaggio perché si può simulare in maniera efficiente la cinetica del ripiegamento della proteina verso la struttura nativa. Tuttavia è uno svantaggio perché non rappresenta perfettamente il sistema reale ma solo in prima approssimazione. Nella realtà infatti l'energia libera ha molti minimi locali e stati metastabili che si osservano durante il processo di ripiegamento. Per avere una maggiore accuratezza, si cercano quindi dei modelli indipendenti dalla struttura di riferimento che introducano frustrazione atta a generare dinamiche di ripiegamento più realistiche, grazie all'uso di termini di potenziale più sofisticati.

3.3.3 Inversione di Boltzmann

Per ottenere l'accuratezza strutturale necessaria per descrivere in maniera generale le proprietà dinamiche di una proteina utilizzando un modello CG, dunque la parametrizzazione non si può basare su un'unica struttura. Si utilizza quindi il metodo dell'inversione di Boltzmann [55], che fornisce la parametrizzazione dall'analisi statistica di grandi insiemi di strutture, ottenute o sperimentalmente o da simulazioni all-atom.

Il potenziale di forza media PMF, o potenziale statistico, calcolabile da un insieme statistico di strutture, è definito come:

$$U(Q_i) = -KT \ln\left(\frac{P(Q_i)}{P_0(Q_i)}\right) \quad (3.23)$$

dove $P_0(Q_i)$ è la distribuzione di probabilità di una variabile interna in uno stato di riferimento.

Si assuma che l'energia totale di un sistema possa essere scomposta in termini scorrelati tra loro, ognuno dipendente da una variabile CG:

$$U(Q) = \sum_i U_i(Q_i) \quad (3.24)$$

dove $U_i = U^{bond}, U^{angle}, \dots$, $Q_i = r_{i,i+1}, \alpha_i, \dots$ e Q è l'intero set di coordinate. Assumendo che le variabili siano completamente indipendenti tra loro, le loro funzioni di probabilità saranno moltiplicative, così la distribuzione di probabilità di una variabile CG è data da:

$$P(Q_i) \propto \int dQ_1, \dots, dQ_{i-1}, dQ_{i+1}, \dots, dQ_n \exp\left(\frac{-U(Q)}{KT}\right) = \exp\left(\frac{-U(Q_i)}{KT}\right) \quad (3.25)$$

Si ottiene così la formula dell'inversione di Boltzmann:

$$U(Q_i) = -KT \ln P(Q_i) + cost. \quad (3.26)$$

Considerando un sistema di particelle non interagenti tra loro e distribuite in modo random in uno spazio tridimensionale, $P_0(Q_i)$ è solitamente nota e PMF è in prima approssimazione uguale all'inversione di Boltzmann. In questo modo si possono ricavare i termini del campo di forze dalla formula dell'inversione di Boltzmann.

3.4 Studi di simulazione della dinamica della proteasi

Molti studi di simulazione e modellistica a vari livelli della proteasi di HIV esistono in letteratura e sono riassunti in tabella 3.1. Qui ci si interessa in particolare dei modelli coarse grained.

Il modello minimalista accoppiato alla parametrizzazione basata sull'inversione di Boltzmann è stato in precedenza applicato alla proteasi di HIV, per comprendere meglio il processo di legame della proteasi con il substrato e la dinamica dei flap [56], raggiungendo una scala temporale dei microsecondi per poter analizzare la dinamica dei flap. Sono così stati osservati sia eventi di completa apertura dei flap, dove la distanza tra le punte dei flap passa da 5 Å a 20–30 Å, sia eventi di chiusura. Inoltre sono state viste aperture dei flap che raggiungevano i 40 Å di distanza. Per raggiungere le configurazioni aperte, i flap si spostano lateralmente rispetto al sito attivo, lasciandolo completamente esposto al solvente. Grazie alla lunghezza della simulazione, dell'ordine delle decine di microsecondi, si è visto che gli eventi di apertura si raggruppano in 'bunches' in cui i singoli eventi sono distanti 5–10 ns e questi raggruppamenti sono distanziati da intervalli di circa 100 ns (vedi figura 3.2). Questa particolare dinamica è stata proposta essere la chiave di come la proteasi riconosca il substrato e lo aggiusti nel sito attivo.

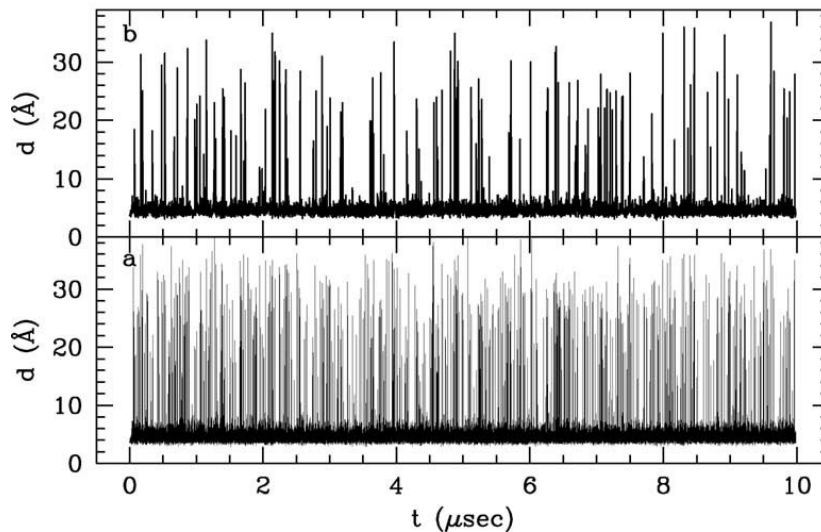


Figura 3.2: Distanza tra le punte dei flap in funzione del tempo: (a) simulazione di $10\mu s$; (b) la stessa aumentando la frequenza di campionamento a $\sim 1ns$ [56].

Nell'articolo di Toth e Borics [57] sono stati proposti due modelli a due fasi del meccanismo di legame tra la proteasi ed il substrato. Nel primo modello, il ligando forma un

Riferimento	modello	risultati
Tozzini e McCammon [56] 2005	one bead CG	dinamica dei flap
Toth e Borics [57] 2006	MM e dinamica di Langevin	il ligando induce l'apertura dei flap
Toth e Borics [25] 2006	MM e dinamica di Langevin	chiusura asimmetrica dei flap
Tozzini et al. [58] 2007	one bead CG e dinamica stocastica	dinamica dei flap in scale di tempo delle decine di microsecondi
Chang et al. [23] 2007	one bead CG in dinamica browniana e all-atom in solvente implicito	l'inibitore <i>XK263</i> può legarsi nel sito attivo mentre i flap sono quasi chiusi
Trylska et al. [26] 2007	one bead CG esteso alle iterazioni proteasi-substrato	la presenza del substrato modifica la mobilità interna della proteasi in modo da favorirne la cattura
Carloni et al. [24] 2009	MD atomistica in acqua esplicita	dinamica dei flap dell'enzima non legato osservandone gli eventi di apertura e chiusura
Soares et al. [59] 2010	simulazioni MM con solvatazione implicita	una singola mutazione ha effetti distinti sui differenti sottotipi di HIV-1
Ramos et al. [60] 2010	MM con tempo scala dei microsecondi	il riconoscimento non è regolato solamente per affinità con il sito attivo
Levy et al. [27] 2011	multiscala	grande apertura dei flap è un fattore importante nella cinetica di legame dei ligandi
Senapati e Karthik [22] 2011	multiscala	modello strutturale per il ciclo catalitico della proteasi
Schiffer et al. [61] 2012	multiscala	il riarrangiamento del nucleo idrofobico è essenziale per l'attività enzimatica della proteasi

Tabella 3.1: Lista di lavori riguardanti la dinamica della proteasi di HIV-1

complesso con la proteasi di HIV-1 entrando nel sito attivo della proteasi quando ha i flap aperti e inducendo successivamente la chiusura dei flap. Nel secondo modello, la vicinanza del ligando induce l'interruzione delle interazioni tra i flap appena il ligando si avvicina alla PR in conformazione dei flap semi-aperti. Questo fa sì che i flap adottino conformazioni aperte quando il ligando entra nel sito attivo della proteasi. Il passo successivo di questo secondo modello è la chiusura dei flap indotta dalla presenza del ligando nel sito attivo. Entrambi questi modelli sono in due step se nel secondo modello l'apertura dei flap e l'accesso del ligando nel sito attivo sono considerati come un singolo step. I risultati ottenuti supportano entrambi i modelli di simulazione.

I tempi di apertura dei flap in questo lavoro differiscono dai risultati ottenuti in simulazioni precedenti. Questo può essere attribuito alla differenza della temperatura di simulazione, e/o alla modellizzazione della viscosità del sistema. Studi precedenti modellavano l'ambiente acquoso utilizzando un modello esplicito d'acqua, mentre in questo studio è stata usata una solvatazione continua con impostazione a bassa viscosità. È stato dimostrato che a basse viscosità aumenta la velocità di folding delle proteine e aumenta il campionamento conformazionale. Pertanto, l'uso in questo studio di solvatazione continua e a bassa viscosità ha consentito un campionamento maggiore della proteasi rispetto ad una simulazione in acqua esplicita nella stessa scala dei tempi, impiegando un minor di tempo di calcolo.

In un altro lavoro sempre di Toth e Borics [25], sono state eseguite due simulazioni di dinamica molecolare di un complesso che consisteva di un peptide ($CA - p2$) posto nel sito attivo della proteasi di HIV-1 con la conformazione dei flap aperti. $CA - p2$ è un peptide di dieci residui (Lys-Ala-Arg-Val-Leu-Ala-Glu-Ala-Met-Ser) derivato dalla poliproteina Gag-Pol di HIV-1 e contiene un sito di taglio della proteasi. Durante la simulazione il substrato induceva la chiusura dei flap in modo asimmetrico attraverso un gruppo di stati idrofobici intermedi. La natura dei residui della proteasi è importante per il processo di chiusura dei flap ed è conservata. Nel corso della simulazione, il raggio di girazione del complesso diminuisce gradualmente, consistentemente con l'osservazione che la struttura del complesso con il substrato legato alla proteasi è più compatta riguardo al complesso con il substrato libero.

I contatti iniziali tra i flap e il substrato erano stabiliti attraverso i legami a idrogeno tra i residui *Gly48* dei flap e il substrato. Questo è consistente con l'ipotesi che la proteasi riconosca una particolare forma piuttosto che la sequenza di un substrato e che il substrato debba avere una conformazione estesa per prendere parte alle numerazioni interazioni a idrogeno con la proteasi necessarie per legarsi ad essa.

Il modello del campo di forze proposto in [56] è stato poi usato per studiare la dinamica della proteasi in una scala di tempo delle decine di microsecondi in vari ensemble statistici, utilizzando il termostato di Berendsen, la dinamica di Langevin e la dinamica browniana [58]. È stato trovato che le strutture chiuse ricavate dalle simulazioni erano in buon accordo con la struttura cristallografica di riferimento che è in conformazione chiusa (codice PDB: 1HHP) e con una struttura cristallografica in conformazione semi-aperta (codice PDB: 1TW7), tranne che nella regione delle punte dei flap, che sono soggette a molte fluttuazioni come visto precedentemente, confermando che il modello riproduce correttamente le strutture nei passi intermedi dell'apertura dei flap. Inoltre questo modello predice l'esistenza di conformazioni aperte ed ampiamente aperte, che probabilmente non possono essere osservate in una forma cristallizzata della proteasi.

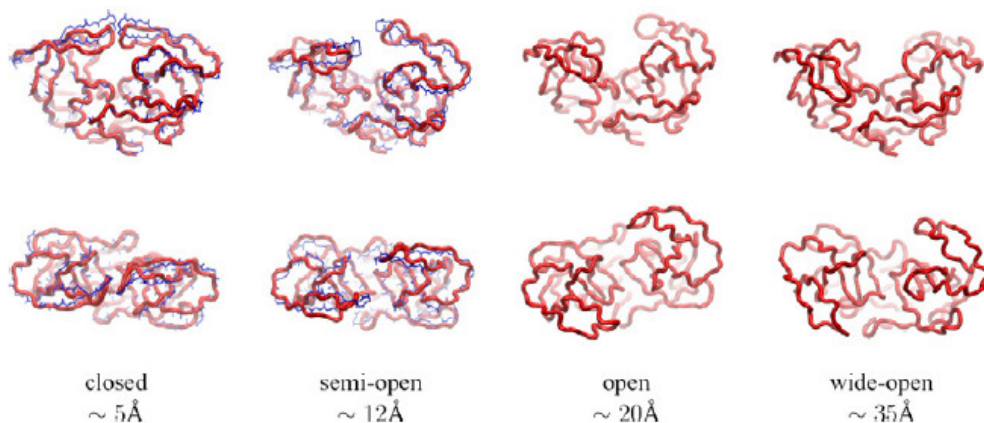


Figura 3.3: Stati conformazionali della proteasi. Le strutture rosse sono prese dalla simulazione NVT, quelle blu sono strutture cristallografiche blue (codice PDB: 1HHP e 1TW7 rispettivamente per le conformazioni chiusa e semi-aperta). E' indicata la distanza tra le punte dei flap (Ile50-Ile149) [58].

In uno studio successivo [23], è stato applicato il modello CG [56, 58] e l'algoritmo di dinamica browniana (BD) per simulare il legame con un oligopeptide modello del substrato naturale e del piccolo inibitore *XK263*, combinato con una simulazione di dinamica molecolare in una rappresentazione atomistica per il sistema *XK263* – *proteasi*.

I risultati di queste simulazioni hanno mostrato che i flap devono aprirsi per legare il substrato peptidico, e quando è nel sito attivo i flap si chiudono consentendo all'enzima di funzionare. Se i flap sono chiusi, il substrato esplora i dintorni dell'enzima fino all'apertura dei flap, quando si lega al sito attivo. In quest'ultimo caso, il substrato può formare un'interazione intermolecolare tra uno dei flap che può causare un movimento asimmetrico dei flap, indebolendo l'attrazione tra le punte e quindi facilitandone l'apertura. Inoltre i flap possono avere conformazioni completamente aperte e chiuse quando un ligando non è vicino al sito di legame.

Tuttavia un inibitore più piccolo come *XK263* può legarsi nel sito attivo anche mentre i flap sono quasi chiusi. Il percorso più comune viene infatti dal una piccola apertura in un lato della proteasi, nonostante l'inibitore impieghi in questo caso più tempo per il processo di legame. Infatti a volte, prima che si chiudano i flap, *XK263* può lasciare il sito di legame, con conseguente insuccesso nel binding. Per il ligando peptidico invece non è possibile accedere al sito di legame quando i flap sono in conformazione semi-chiusa.

La durata della vita media delle conformazioni aperte e chiuse dei flap erano rispettivamente di 70 e 430ns. A differenza della proteasi non legata con il substrato, quando il substrato è nel sito attivo ma i flap sono ancora aperti, essi rimangono in questo stato molto più a lungo, impiegando circa 1μs per chiudersi. Analogamente, quando il substrato è presente nel sito attivo con i flap chiusi, i flap rimangono chiusi stabilmente almeno fino a che non avviene l'idrolisi. Il substrato interagisce con la proteina, e quindi altera il movimento interno della proteasi. Sono state anche osservate interazioni *XK263*-proteasi con la dinamica dei flap, ma questi effetti sono più piccoli di quelli trovati nel legame substrato-proteasi.

Nelle due simulazioni di dinamica molecolare atomistiche [23], la dinamica molecolare suggerisce che con conformazioni adeguate, *XK263* può entrare nel sito di legame senza

aprire i flap. Ciò è in buon accordo con le simulazioni CG. La seconda simulazione di dinamica molecolare mostra un altro possibile percorso di legame, in cui *XK263* può indurre un'apertura parziale dei flap e legarsi alla proteasi con i flap non completamente aperti. Inoltre le simulazioni CG hanno mostrato che dopo che il ligando peptidico si sposta nel sito attivo della proteasi, la proteasi ha bisogno di più tempo per la chiusura dei flap, rispetto a quando c'è l'inibitore *XK263*. Questa osservazione è coerente con i risultati delle simulazioni di dinamica molecolare atomistica.

Successivamente il modello sviluppato in [23, 56, 58] è stato usato in [26] per il processo di legame tra un substrato peptidico e la proteasi di HIV-1 e la dissociazione dei prodotti dopo l'idrolisi del substrato, tramite l'estensione dei campi di forza del precedente modello CG. Le simulazioni mostrano che i flap hanno bisogno di aprirsi per legarsi con il peptide e che l'interazione della proteasi con il substrato influenza la frequenza e l'intervallo di apertura dei flap, stabilizzando leggermente la conformazione aperta rispetto al caso della proteasi non legata. Il rilascio dei prodotti, che avviene nel timescale dei nanosecondi, non richiede l'apertura dei flap, perché possono scivolare fuori dalla fessura ai lati dell'enzima.

Le analisi delle fluttuazioni del substrato nel sito attivo della proteasi hanno mostrato che il moto del substrato è correlato con il movimento del sito di taglio, con i flap e soprattutto con le punte dei flap.

Dopo la diffusione iniziale verso l'enzima, il substrato attende in prossimità del sito di legame. Quando il substrato entra nella fessura di legame, si ha la chiusura di un flap. Dopo che il substrato si accomoda correttamente nella fenditura, l'altro flap si chiude. Nella figura 3.4 abbiamo una rappresentazione del processo. Le fluttuazioni delle punte dei flap diminuiscono rispetto a quelle all'inizio della simulazione, quando il substrato era fuori dalla proteasi.

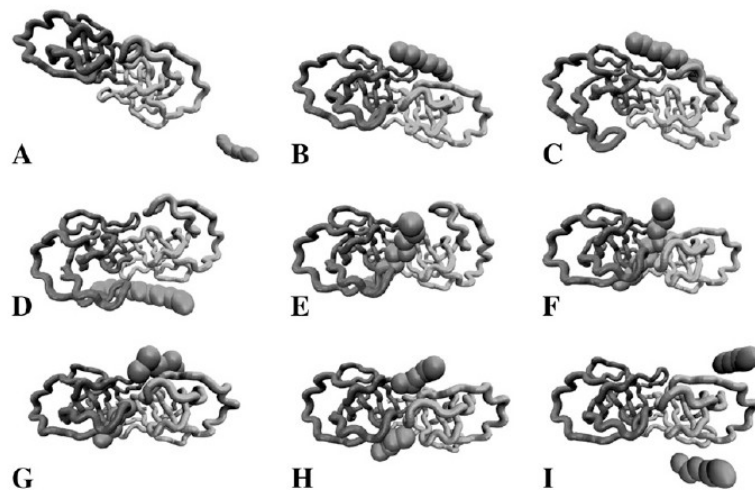


Figura 3.4: Serie di immagini del processo di legame, taglio e rilascio di un substrato ad opera della proteasi. (A) Diffusione iniziale intorno alla proteasi (≈ 20 ns). (B) Il substrato aspetta l'apertura dei flap della proteasi. (C) Apertura dei flap (≈ 80 ns) ma il substrato non è correttamente posizionato. (D) Il substrato si sposta dall'altro lato del sito di taglio (≈ 120 ns) e i flap si aprono nuovamente (≈ 140 ns). (E) Il substrato entra nel sito di taglio (≈ 215 ns) e un flap si chiude sopra il substrato. (F) Corretto posizionamento del substrato con entrambi i flap chiusi (≈ 250 ns). (G) Rottura del substrato indotta a 400 ns. (H) Rilascio dei prodotti e (I) loro diffusione [26].

I risultati suggeriscono che la presenza del substrato modifica la mobilità interna della proteasi in modo da favorirne la cattura.

La dinamica molecolare con solvente esplicito è usata in [24] per simulare la dinamica dei flap dell'enzima non legato osservando l'apertura spontanea e gli eventi di chiusura. Fino a questo momento l'elevato costo computazionale ha limitato la scala temporale sulle simulazioni MD a $1.6 \mu s$ con solvente esplicito. Nel lavoro si focalizza sul substrato Thr-Ile-Met-Met-Gln-Arg (SUB), sito di taglio p2-NC della poliproteina virale gag-pol. La scelta di questo ligando è motivata dal fatto che la maggior parte degli inibitori nell'uso clinico sono peptidomimetici, quindi si basano su un substrato naturale e potrebbero avere diverse caratteristiche in comune. La via principale di legame del piccolo substrato non comporta una completa apertura dei flap, come notato in precedenti lavori. La simulazione è stata inizializzata con il substrato completamente al di fuori della cavità e orientato in posizione sfavorevole. Durante la simulazione, sono stati osservati quattro eventi di legame e dieci di non legame. Dall'analisi della traiettoria, i percorsi più probabili di legame/non legame erano quelli con minor energia libera e comprendevano i seguenti stati (vedi figura 3.5):

- **B4.** SUB è solvatato senza alcun contatto con l'enzima; i flap sono piuttosto chiusi.
- **TS.** SUB esterno alla cavità, perpendicolare rispetto all'orientamento nel complesso, con *Thr(P3)* vicino alla cavità.
- **B3.** La cavità, allargata da un moderato spostamento dei flap, consente al SUB di entrare.
- **B2.** Le molecole d'acqua di collegamento tra SUB e le punte dei flap vengono espulse, stringendo la cavità.
- **B1.** Le molecole d'acqua di collegamento tra SUB e la diade catalitico sono espulse. Viene introdotta una molecola d'acqua (W301) fra SUB e le punte dei flap.

Lo stato più popoloso (46,9%) è il complesso B1. Questo risultato è una previsione della struttura sperimentale e della stabilità del complesso. La popolazione di equilibrio del cluster corrispondente all'enzima separato dal ligando (B4) è invece praticamente trascurabile.

Per un piccolo substrato si è trovato che nei percorsi più probabili di associazione e di dissociazione i flap della proteasi non si aprono ampiamente, e il substrato entra dentro la cavità dell'enzima dallo stretto canale laterale, in accordo con gli studi precedenti. Quando il ligando si avvicina all'enzima nelle prime fasi del legame e nello stato di transizione, forma legami a idrogeno con i residui Asp30', Ile47', Gly48', and Glu35 che sorgono intorno all'apertura del canale che conduce alla cavità catalitica.

Si è anche notato che i flap totalmente aperti hanno una popolazione molto bassa.

In [59] viene analizzato il legame dell'inibitore nelfinavir (NFV) e del sito di taglio del substrato CA/p2 di Gag con la proteasi dei sottotipi B e C di HIV-1 attraverso simulazioni MM. Sono stati investigati sia il tipo naturale (*wild-type*) vari mutanti resistenti ai farmaci, insieme alle mutazioni compensatorie N83T e N88D, osservate in vitro e in vivo quando il sottotipo C acquisisce la mutazione D30N. D30N sembra facilitare cambiamenti conformazionali nella proteasi del sottotipo B ma non in quella del sottotipo C, e questo potrebbe essere

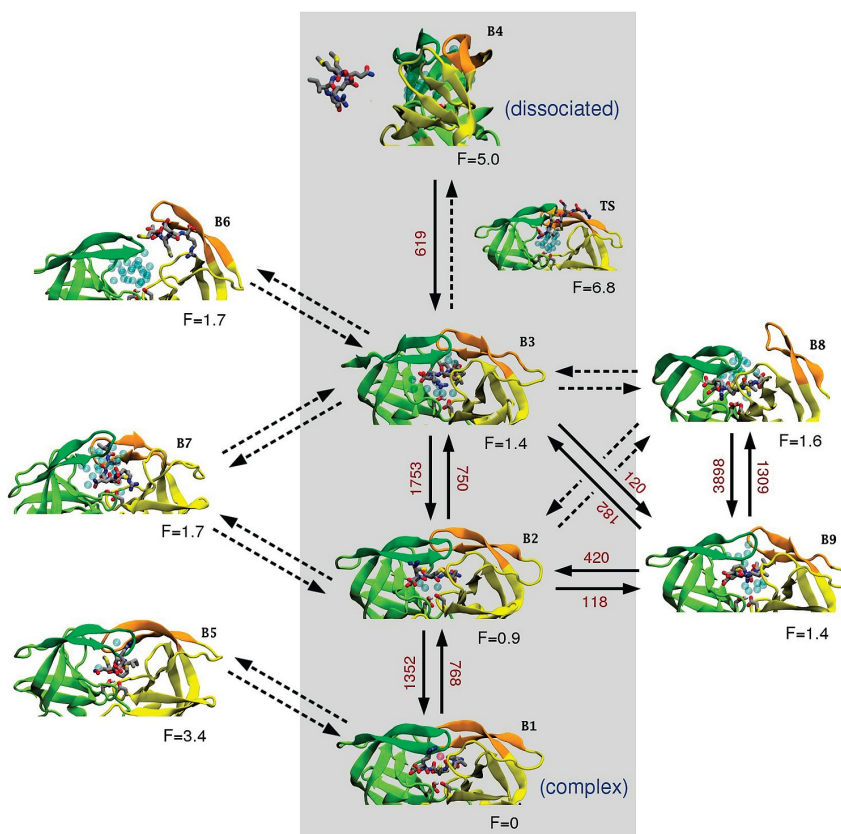


Figura 3.5: Strutture rappresentative coinvolte nel principale meccanismo di legame/non legame (nella parte colorata). Sono anche mostrati gli stati $B6 - B9$, che non sono coinvolti nel percorso principale. Le molecole d'acqua dentro la cavità dell'enzima sono mostrate come sfere blu, tranne W301 che è colorata di rosso. Sotto ogni struttura è riportata l'energia libera dello stato in kcal/mol. Le frecce sono etichettate con i corrispondenti rate di transizione (in ms^{-1}), le frecce tratteggiate indicano transizioni con un rate minore di $100 ms^{-1}$ [24].

associato ad una separazione di un'elica α della proteasi. Inoltre le aree totali di contatto di NFV o del substrato CA/p2 con la proteasi mutante sono collegate a cambiamenti nella stabilità dell'enzima. Questo suggerisce che i diversi meccanismi molecolari contribuiscono alla resistenza ai farmaci, e viene proposto che una singola mutazione ha effetti distinti sui differenti sottotipi di HIV-1. Le energie di legame calcolate per NFV mostrano che la mutazione D30N indebolisce le interazioni tra la proteasi e l'inibitore per entrambi i sottotipi di HIV-1 studiati.

Le energie di legame calcolate per i complessi proteasi-CA/p2 mostrano invece che D30N causa interazioni più favorevoli tra la proteasi e il substrato CA/p2 e rivelavano che tutti i mutanti presentavano interazioni più forti se paragonate ai tipi naturali.

Sono in seguito state eseguite simulazioni MM con un tempo scala dei microsecondi scoprendo che alcune sequenze della poliproteina Gag-Pol, che non vengono scisse (non substrates), hanno una maggiore affinità con il sito attivo della proteasi (PR) di HIV-1 rispetto a un substrato, quindi il riconoscimento non è regolato solamente per affinità con il sito attivo [60]. La proteasi è altamente specifica, cioè taglia solo un insieme molto specifico di sequenze amminoacidiche.

Dall'analisi dei risultati e dei dati sperimentali, è stato proposto che il riconoscimento si basa sulla specificità geometrica del complesso multiproteico $PR : Gag$ e $PR : Gag - Pol$, che seleziona i residui che si trovano nella posizione specifica che li rende accessibile al sito attivo per il taglio. Così per verificare questa affinità è stata calcolata la differenza di energia libera relativa nel legame di un substrato rispetto ad alcuni nonsubstrati:

$$\Delta\Delta G_{bind}^{S \rightarrow nS} = \Delta G_{bind}^{nS} - G_{bind}^S = \Delta G_P^{S \rightarrow nS} - G_{solv}^{S \rightarrow nS}$$

dove $G_{solv}^{S \rightarrow nS}$ e $\Delta G_P^{S \rightarrow nS}$ sono le energie libere non fisiche di trasmutazione dal substrato al nonsubstrato in soluzione e legato alla proteasi, rispettivamente. Il calcolo richiedeva la trasformazione virtuale del substrato nel nonsubstrato, che coinvolgeva il mutare un numero molto elevato di amminoacidi. I risultati ottenuti prevedono che il nonsubstrato TKALTEVI ha una maggiore affinità alla proteasi rispetto al substrato ARVLAEAM, anche se la proteasi non taglia il primo ma taglia sempre il secondo. Pertanto, la selettività specifica per il substrato non può essere basata esclusivamente sull'affinità con il sito attivo.

La proposta che la specificità della proteasi è determinata dalla capacità delle catene laterali degli amminoacidi dei substrati di legarsi agli otto sottositi all'interno dell'enzima è incompleta. La proteasi può legare una grande varietà di peptidi, e alcuni studi hanno mostrato che:

- una varietà di residui amminoacidici può essere accolta in ciascuno dei sottositi dell'enzima;
- i sottositi sono in grado di agire in maniera indipendente nel riconoscimento degli amminoacidi nella posizione corrispondente;
- nei substrati naturali, amminoacidi con differenti proprietà e dimensioni delle catene laterali possono occupare la stessa posizione nella tasca di legame della proteasi.

Non ci sono proposte che possono spiegare la specificità della proteasi basata solo sulle interazioni dei substrati all'interno della tasca di legame, perché parecchi nonsubstrati possono stabilire interazioni più forti.

I dati sperimentali mostrano che il processo di idrolisi attuato dalla proteasi dipende dal folding della poliproteina ed un sito di taglio dipende dal ripiegamento della poliproteina. Da questi risultati si evince che un modello rigido chiave-serratura non è appropriato per descrivere il riconoscimento del substrato. Così viene proposta una nuova ipotesi per spiegare il riconoscimento del substrato da parte della proteasi. La proteasi scinde solo i substrati che sono sufficientemente esposti per un dato tempo, quindi molti nonsubstrati non vengono scissi pur avendo una maggiore affinità, perché sono sepolti all'interno della poliproteina e fisicamente nascosti alla proteasi. Tra i residui esposti, il riconoscimento si basa sulla specificità geometrica dei complessi $PR : Gag$ e $PR : Gag - Pol$ che determina quali residui si trovano in una posizione accessibile al sito attivo.

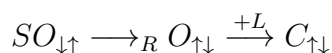
Per superare la limitazione della scala dei tempi delle simulazioni di dinamica molecolare convenzionale, il metodo utilizzato da Levy et al. [27] cerca di studiare la dipendenza della temperatura con i collegamenti degli stati della proteasi. Alle basse temperature l'ampia apertura dei flap è dominata da un piccolo numero di percorsi mentre a temperature elevate

la transizione avviene attraverso numerose vie strutturalmente eterogenee. A $T = 285K$, il tempo di rilassamento per la transizione semi-open \rightarrow closed è risultato di $\tau_{sc} \approx 33ns$. Per la transizione semi-aperto \rightarrow completamente aperto, il tempo di rilassamento è $\tau_{so} \approx 375ns$. I risultati rivelano un quadro cinetico con veloci riarrangiamenti interni e aperture infrequenti. Questo è qualitativamente consistente con le dinamiche dei flap osservate negli studi NMR, che ha rivelato movimenti su due scale temporali. Le scale temporali calcolate in questo studio sono anche approssimativamente paragonabili ai risultati delle simulazioni MD del range da 40 a 300ns a temperatura ambiente.

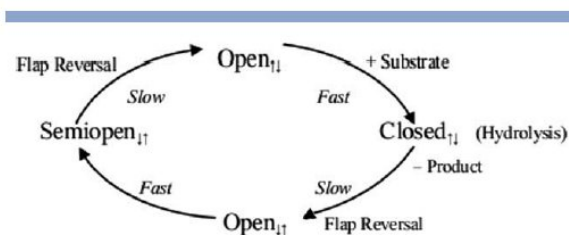
Queste analisi sembrano suggerire che l'ampia apertura dei flap è un fattore importante nella cinetica di legame dei ligandi della proteasi di HIV-1. In [22] si cercano le diverse conformazioni della proteasi durante il suo ciclo catalitico e propongono un modello strutturale per il ciclo catalitico della proteasi di HIV-1. Nelle conformazioni semiaperte e chiuse dei flap c'è una disposizione opposta dei flap.

Le analisi indicano che né la conformazione semiaperta né una conformazione aperta con l'ordinamento dei flap da sotto a sopra ($\downarrow\uparrow$) può legare il substrato e raggiungere una conformazione chiusa per l'idrolisi. Questo è dovuto dall'ingombro sterico dei flap causato dall'allineamento caratteristico dei residui. I flap necessitano di invertire il loro ordinamento visitando una conformazione aperta. L'evoluzione di una conformazione aperta da uno stato semiaperto è un processo lento, che da studi NMR è stato trovato dell'ordine dei microsecondi, mentre la transizione da uno stato aperto a uno chiuso dopo il legame con un substrato è un processo più veloce, in una scala temporale dei nanosecondi da quanto si evince da studi NMR e da simulazioni di dinamica molecolare.

Lo studio suggerisce che avviene il seguente fenomeno di riordinamento dei flap durante il processo di legame del substrato:



dove R sta per inversione dell'ordinamento dei flap. Dopo l'uscita dei prodotti dell'idrolisi, l'enzima deve tornare da una conformazione chiusa a una semiaperta, le quali hanno un ordinamento opposto dei flap da come si può vedere in figura 3.6.

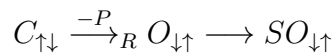


Scheme 1

HIV-1 protease flaps adopt unique ordering at different stages in the catalytic cycle.

Figura 3.6: Ordinamento dei flap della proteasi durante il ciclo di HIV-1 [22]

Nell'articolo si evidenzia che il fenomeno di riordinamento dei flap prende luogo durante il processo di rilascio dei prodotti:



dove $-P$ sta per il rilascio dei prodotti, O per configurazione aperta, C per chiusa e SO per semi-aperta.

In [61] viene mostrato che il riarrangiamento del nucleo idrofobico è essenziale per l'attività enzimatica della proteasi, in quanto i risultati ottenuti supportano che la flessibilità del nucleo idrofobico contribuisca all'efficienza del taglio del substrato.

Il processo di legame del ligando nella proteasi comporta cambiamenti conformazionali nei flap della proteasi e nei residui del nucleo idrofobico. Simulazioni di dinamica molecolare hanno suggerito che i cambiamenti conformazionali che si verificano durante la riorganizzazione delle catene laterali nel core idrofobico siano un possibile meccanismo che colpisce la sensibilità ai farmaci.

3.5 Studio di simulazione sulla poliproteina Gag

Per quanto riguarda gli studi di simulazione sulla poliproteina Gag poco è presente in letteratura. Un lavoro interessante è quello di Gregory Voth e Gary Ayton [62]. In questo lavoro sono state fatte delle simulazioni multiscala sviluppate tramite una combinazione di dati sperimentali e di metodi coarse-grained, al fine di esplorare la struttura del virione immaturo di HIV-1. Infatti per sviluppare il modello CG della Gag sono state prese le strutture atomiche dei domini che compongono la poliproteina (MA , CA_{NTD} , CA_{CTD} , NC e $SP1$), usando come siti del modello CG i vari atomi di carbonio- α . Nel modello CG sono stati creati $N_D = 16$ domini, 14 domini rappresentanti la Gag più due rappresentanti la membrana lipidica, contenenti ognuno N_α siti CG. I domini Gag 1 – 14 sono reti elastiche connesse da collegamenti flessibili.

Ci sono γ tipi di interazioni di coppia tra due siti e l'energia totale del sistema Gag-membrana è data da:

$$U = \frac{1}{2} \sum_{\alpha=1}^{N_D} \sum_{i_\alpha=1}^{N_\alpha} \sum_{\beta=1}^{N_D} \sum_{j_\beta=1}^{N_\beta} \sum_{\gamma=1}^{N_D} \zeta_\gamma^{\alpha\beta} u_{i_\alpha j_\beta}^\gamma \quad (3.27)$$

dove $\zeta_\gamma^{\alpha\beta}$ è un'interazione di coppia di tipo γ e $u_{i_\alpha j_\beta}^\gamma$ è una funzione con valore 0 o 1 a seconda di dove sia richiesta l'interazione tra la coppia di siti i_α i_β , rispettivamente nel dominio α e β .

Dalle simulazioni si è visto che solo alcuni siti CG nel dominio CA_{CTD} sono responsabili del mantenimento della simmetria esagonale di p6 nel virione immaturo. Inoltre per stabilizzare questa regione le forze di interazione tra i siti CG di CA_{CTD} devono essere in un intervallo tra 3 – 4 kcal/mole.

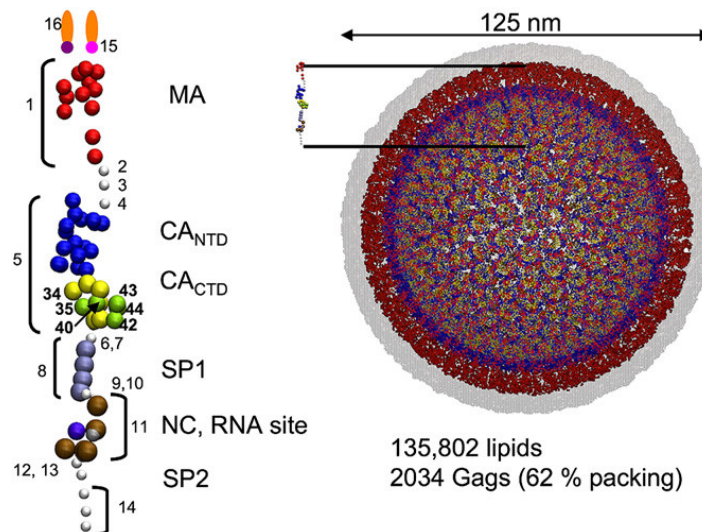


Figura 3.7: Modello CG di Gag/membrana. Nella parte sinistra è mostrata la composizione dei domini. A destra la struttura iniziale del virione: 2034 poliproteine Gag, in fasci di sei, in un virione di 125 nm [62].

3.6 Sommario e motivazione del lavoro di tesi

Dall'analisi dei dati di simulazione e sperimentali emerge il seguente quadro:

1. Per quanto riguarda la dinamica della proteasi in assenza di substrato, le simulazioni a vari livelli indicano che gli eventi di apertura dei flap sono sporadici e avvengono sulla scala del micro-millisecondo, che è piuttosto grande per avere un campionamento estensivo delle configurazioni con simulazioni atomistiche. Esiste tuttavia un modello minimalista che mostra di riprodurre bene gli eventi di apertura e la dinamica interna con un costo computazionale molto modesto.
2. Per quanto riguarda le interazioni della proteasi con modelli del substrato naturale e con vari inibitori, i dati sperimentali indicano quali sono le sequenze peptidiche tagliate con maggiore frequenza, anche se ci sono evidenze secondo le quali il meccanismo di riconoscimento del substrato da parte della proteasi coinvolge anche, e forse principalmente, la struttura del peptide, oltre che la sequenza. Simulazioni a vari livelli indicano che mentre piccoli inibitori possono entrare nel sito attivo anche con la proteasi chiusa, per agganciare il substrato peptidico essa deve aprirsi; questo processo è ben descritto dal modello minimalista.
3. Solo poche simulazioni dell'intera Gag esistono in letteratura, e ad un livello di risoluzione molto basso.

Ciò considerato, appare opportuno ed utile fare uno studio di simulazione al livello di modello CG minimalista dell'interazione tra proteasi e Gag, che possa indagare alcuni aspetti ancora oscuri del meccanismo di aggancio della proteasi al suo substrato rappresentato in maniera integrale. Nel prossimo capitolo verrà illustrato in dettaglio il modello, che utilizza

il campo di forze già ottimizzato della proteasi. Il modello della Gag e dell'interazione Gag-proteasi è invece un contributo originale di questo lavoro di tesi, così come la produzione delle simulazioni dell'interazione e la loro analisi, illustrate nel prossimo capitolo.

Capitolo 4

Interazione tra proteasi di HIV-1 e poliproteina Gag: ottimizzazione dei campi di forze e simulazioni

In questo capitolo sono ottimizzati i campi di forze minimalisti per il sistema proteasi-Gag di HIV-1, realizzate le simulazioni e analizzati i risultati. Il campo di forze della proteasi è quello già ottimizzato nelle referenze [23, 26, 56, 58] e non è stato cambiato. Tuttavia, per controllare il funzionamento del modello della proteasi sono state inizialmente eseguite alcune simulazioni per verificarne la dinamica dei flap e le interazioni tra proteasi e substrato. Il campo di forze della Gag e dell'interazione Gag-pr è stato invece costruito e ottimizzato in conformità con quello della proteasi. Sono state realizzate simulazioni riguardanti la Gag isolata, in cui osserva la flessibilità del linker tra i domini MA e CA_{NTD} , e l'interazione tra la poliproteina Gag e la proteasi nel sito di taglio tra MA e CA_{NTD} .

Per eseguire queste simulazioni, è stato utilizzato il software DL_POLY con alcune modifiche, aggiunte per implementare la dinamica di Langevin e le forme funzionali per il campo di forze della Gag.

4.1 Scelta del software DL_POLY

DL_POLY [47] è un pacchetto di software per le simulazioni di dinamica molecolare, cosiddetto general purpose, cioè utilizzabile in linea di principio per qualsiasi tipo di sistema molecolare in fase solida, condensata o cristallina, scritto in FORTRAN90. Nel pacchetto sono inclusi vari campi di forza e condizioni al contorno e sono codificati vari algoritmi di simulazione.

Come discusso precedentemente, il campo di forza è il set di funzioni necessario per definire le interazioni in un sistema molecolare. Queste possono avere un'ampia varietà di forme analitiche. La descrizione del campo di forza assume che il sistema simulato sia descritto come un insieme di oggetti interagenti che possono essere sia atomi che pseudo-atomi nella rappresentazione coarse-grained. DL_POLY implementa una forma molto generale del campo di forze, che è riportato in appendice C. Questo software è stato scelto per la sua flessibilità. Infatti i modelli coarse grained ed in particolare il minimalista necessitano forme

analitiche non standard. Alcune di esse sono già implementate in DL_POLY, quelle mancanti possono essere incluse con semplici modifiche del programma (si veda appendice C).

4.2 Il campo della proteasi

Il campo di forze del modello minimalista della proteasi è la somma dei seguenti termini:

$$U = U_b + U_\theta + U_\phi + U_{nb}^{loc} + U_{nb}^{non-loc}$$

Nel modello qui usato, il termine di legame U_b è una somma di vincoli che mantengono fisse le distanze tra gli atomi di C_α legati lungo la catena amminoacidica. U_θ è un potenziale quartico

$$U_\theta = \frac{k_\alpha}{2}(\theta - \theta_\alpha)^2 + \frac{k'}{3}(\theta - \theta_\alpha)^3 + \frac{k''}{4}(\theta - \theta_\alpha)^4 \quad (4.1)$$

dove θ_α è associato al primo minimo ed il secondo minimo è θ_β [63]. Le relazioni tra $\theta_\alpha, \theta_\beta$ e le costanti k_α e k_β (da cui derivano i coefficienti k' e k'') sono espresse dalle seguenti equazioni:

$$\Delta = \theta_\beta - \theta_\alpha$$

$$k' = -\frac{k_\beta + 2k_\alpha}{\Delta}$$

$$k'' = \frac{k_\alpha + k_\beta}{\Delta^2}$$

Infine le altezze delle barriere di potenziale sono date dalle seguenti relazioni:

$$\Delta U_\theta = \frac{\Delta^2}{12}(k_\beta - k_\alpha)$$

$$\Delta U_\theta^\alpha = \frac{\Delta^2}{12} \frac{k_\alpha^3}{(k_\alpha + k_\beta)^3} (k_\alpha + 2k_\beta)$$

$$\Delta U_\theta^\beta = \frac{\Delta^2}{12} \frac{k_\beta^3}{(k_\alpha + k_\beta)^3} (k_\beta + 2k_\alpha)$$

Per quanto riguarda il termine del potenziale dei diedri, è dato da una funzione coseno dell'angolo diedro ϕ

$$U_\phi = k[1 - \cos(\phi - \phi_0)]$$

con il valore di equilibrio ϕ_0 dipendente dalle strutture secondarie conosciute a priori.

Per le interazioni non-bonded, sia locali che non locali, viene usato un potenziale di tipo Morse

$$U_{nb} = E_0\{[1 - \exp(-k(r - r_0))]^2 - 1\} \quad (4.2)$$

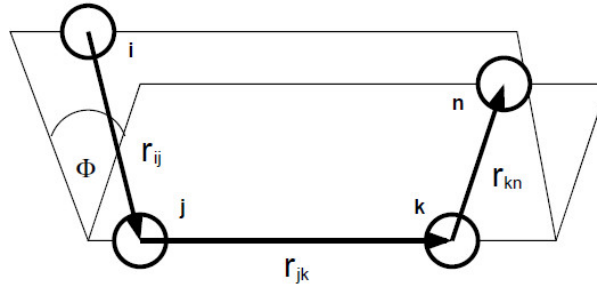


Figura 4.1: Angolo diedro Φ . Le palline rappresentano i quattro atomi di carbonio- α indicati con i, j, k, n necessari per stabilire l'angolo diedro, i simboli r_{ij}, r_{jk} e r_{kn} rappresentano le distanze tra i diversi atomi [47].

in cui r_0 è la distanza di equilibrio, E_0 e k sono la forza ed il range di interazione. Le interazioni locali e non locali sono separate tramite un cutoff di 8\AA . La distanza di equilibrio è settata a 9.5\AA per le interazioni non locali ed è presa da una struttura di riferimento (codice PDB: 1HHP) per le interazioni locali. Per queste interazioni, diminuendo la distanza di equilibrio si ha una maggiore forza del legame, la qual cosa è stata quantificata assumendo una dipendenza empirica tra l'energia di interazione E_0 e la distanza di equilibrio r_0 :

$$E_0 = 6 \exp(-r_0/2.822)$$

Per il terzo parametro del potenziale, è stato assunto il valore fisso di $k = 0.70 \text{\AA}^{-1}$, che è un compromesso tra la necessità di avere un potenziale sufficientemente debole ed un cutoff abbastanza corto per le interazioni non-bonded [58].

I valori dei vari parametri è stato determinato tramite l'inversione di Boltzmann, poiché viene assunto che la distribuzione di probabilità di una coordinata interna, valutata su un insieme statistico di strutture, rappresenti in prima approssimazione il valore dell'energia potenziale della coordinata sotto osservazione.

Tutti i parametri del campo di forze necessari a definire questo modello sono stati fissati come nel lavoro originale [56, 58] e sono riportati in appendice D.

4.2.1 La dinamica dei flap

È stata eseguita una simulazione avente una durata di $10\mu s$ utilizzando la dinamica di Langevin con potenziale dissipativo $\gamma = 2ps^{-1}$ per testare il campo ed avere valori di riferimento. Il passo temporale usato per l'integrazione era $0.04ps$ e la temperatura è stata fissata a 300 K per imitare le condizioni ambientali.

I grafici in figura 4.2 riportano in funzione del tempo i valori della temperatura (a), dell'energia potenziale (b), del potenziale degli angoli (linea rossa in (c)) e dei diedri (linea verde in (c)), e del potenziale dato dai termini non bonded locali (linea rossa in (d)) e dai non bonded non locali (linea verde in (d)).

Come si può vedere dal grafico in figura 4.3, che mostra l'andamento della distanza tra le punte dei flap nel tempo (precisamente tra i C_α degli amminoacidi 50 e 50'), la simulazione mostra eventi di apertura e di chiusura (in figura 4.5 sono mostrate le conformazioni chiusa ed aperta della proteasi). Negli eventi di apertura i flap arrivano a una distanza di separazione

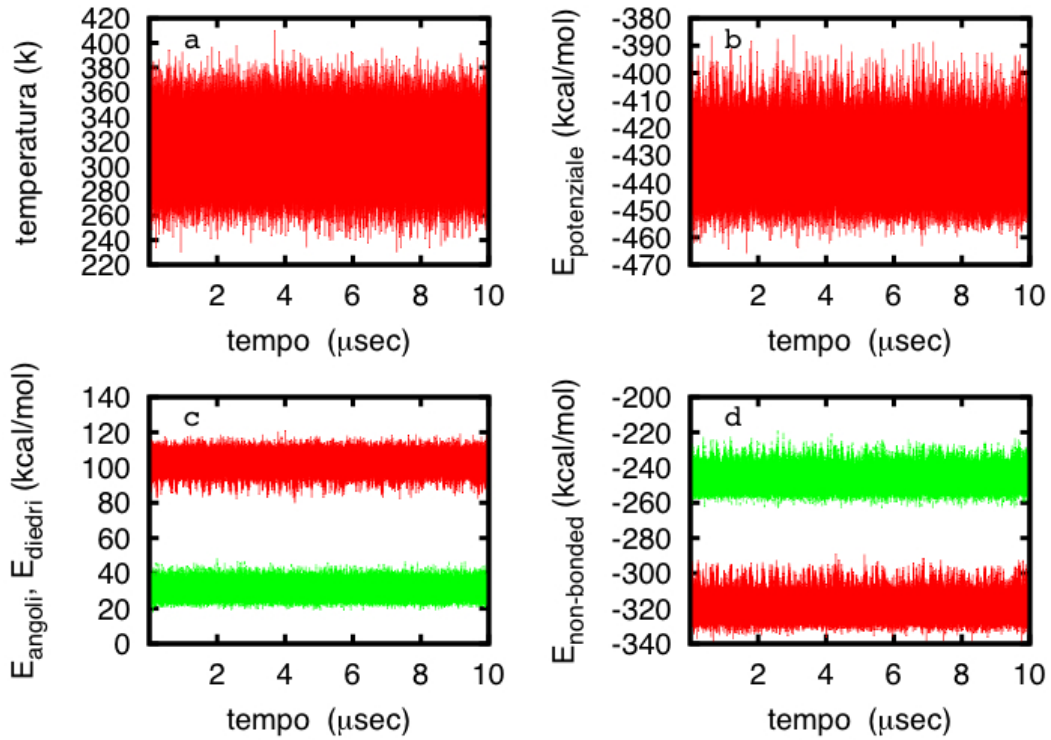


Figura 4.2: Grafici in funzione del tempo della temperatura (a), dell'energia potenziale (b), del potenziale degli angoli (linea rossa in (c)) e dei diedri (linea verde in (c)), e del potenziale dato dai termini non bonded locali (linea rossa in (d)) e dai non bonded non locali (linea verde in (d)) della simulazione con la proteasi isolata.

di $20 - 30\text{\AA}$, e si possono osservare anche eventi di separazione dei flap fino a 40\AA . Durante il fenomeno dell'apertura dei flap, si osserva che essi hanno anche uno spostamento laterale rispetto al sito attivo, il quale viene completamente esposto al solvente.

Date due strutture macromolecolari simili sovrapposte nel modo migliore, S_a e S_b , RMSD è definito come la radice della distanza quadratica media tra gli atomi delle due strutture:

$$RMSD(S_a, S_b) = \sqrt{\frac{1}{N} \sum_{i=1}^N \|R_i^a - R_i^b\|^2} \quad (4.3)$$

con N che rappresenta il numero totale di atomi inclusi nella struttura, mentre R_i^a e R_i^b sono i vettori posizione dell' i -esimo atomo rispettivamente nella struttura S_a ed in quella S_b [65]. Nel grafico di figura 4.4, si osserva il valore dell'RMSD calcolato utilizzando la struttura iniziale come struttura di riferimento, e quindi se ne osservano i cambiamenti in funzione del tempo. Chiaramente, siccome la struttura iniziale di riferimento è chiusa, anche l'RMSD ha valori alti in corrispondenza delle strutture aperte e quindi il suo andamento temporale rispecchia quello della distanza tra le punte dei flap.

È anche utile calcolare la radice delle fluttuazioni quadratiche medie di ogni bead durante la simulazione:

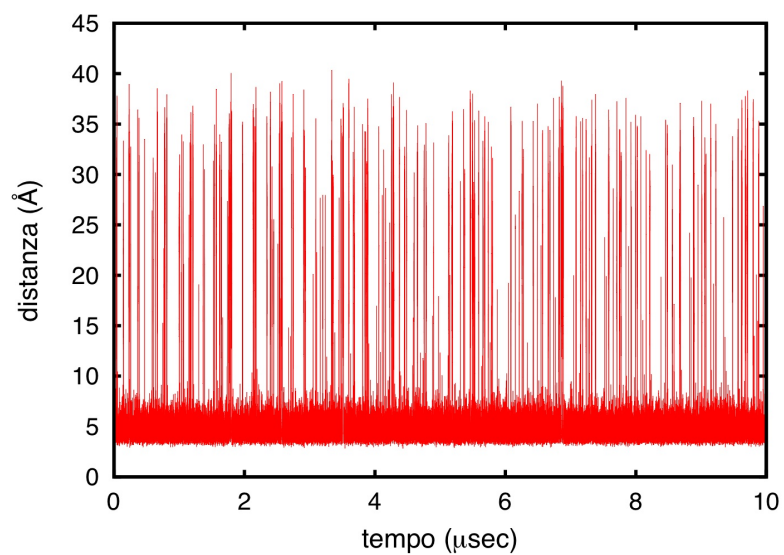


Figura 4.3: Distanza tra le punte dei flap in funzione del tempo nella simulazione con la proteasi isolata. Per distanza tra le punte dei flap si intende la distanza tra gli amminoacidi Ile50 ed Ile50'.

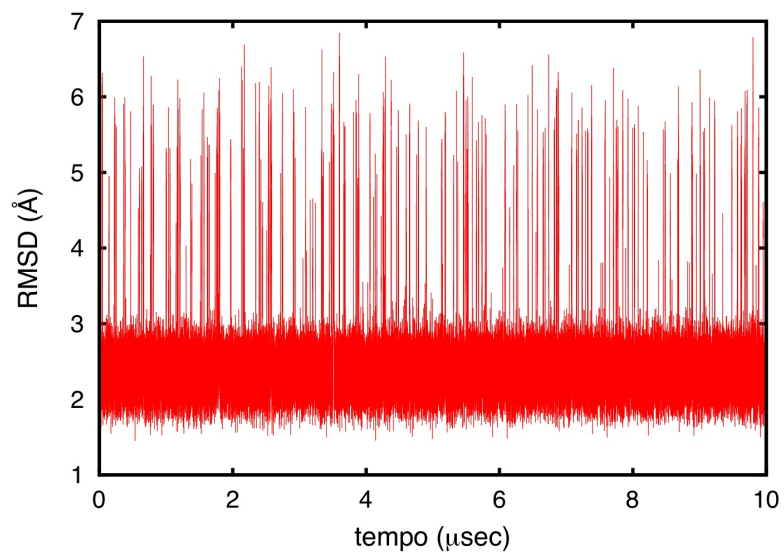
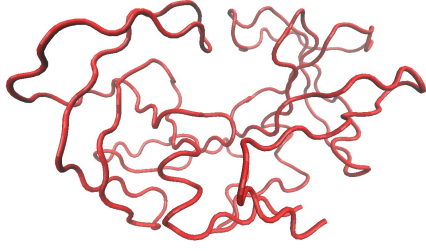
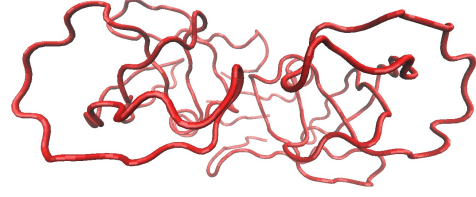


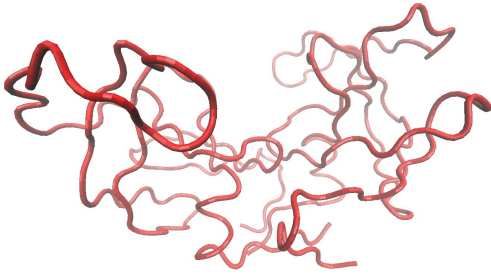
Figura 4.4: RMSD calcolato utilizzando la struttura iniziale come struttura di riferimento.



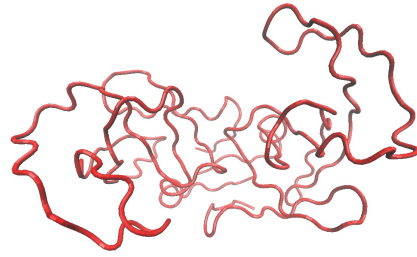
(a) Vista frontale della proteasi in conformazione con i flap chiusi.



(b) Vista dall'alto della proteasi in conformazione con i flap chiusi.



(c) Vista frontale della proteasi in conformazione con i flap aperti.



(d) Vista dall'alto della proteasi in conformazione con i flap aperti.

Figura 4.5: Visualizzazioni della proteasi nelle due conformazioni con i flap aperti e chiusi.

$$RMSF_i = \sqrt{\frac{1}{N} \sum_{t=0}^N (R_i^t - R_i^{ave})^2} \quad (4.4)$$

R_i^t e R_i^{ave} sono le coordinate istantanee e mediate sul tempo dell' i -esimo residuo, in questo caso l'atomo di carbonio- α , e N è il numero di configurazioni istantanee ottenute dalla simulazione. Con RMSF si ha una misura delle fluttuazioni rispetto alla loro posizione media (sia per il calcolo di RMSD che di RMSF, il moto rotazionale e traslazionale globale è stato sottratto allineando le strutture a quella iniziale, in modo da analizzare solo le fluttuazioni interne). RMSF si può confrontare con il fattore di temperatura B incluso nelle determinazioni sperimentali delle strutture cristallografiche che misura l'incertezza sulla posizione. B e RMSF sono legati dalla seguente relazione [65]:

$$RMSF_i^2 = \frac{3B_i}{8\pi^2} \quad (4.5)$$

Si può vedere dalla figura 4.6 che le fluttuazioni sperimentali della proteasi (in rosso) sono abbastanza simili a quelle ottenute per mezzo della simulazione (in verde) tranne che nella zona dei flap, corrispondente ai residui 45-55 (primo monomero) e 144-154 (secondo monomero).

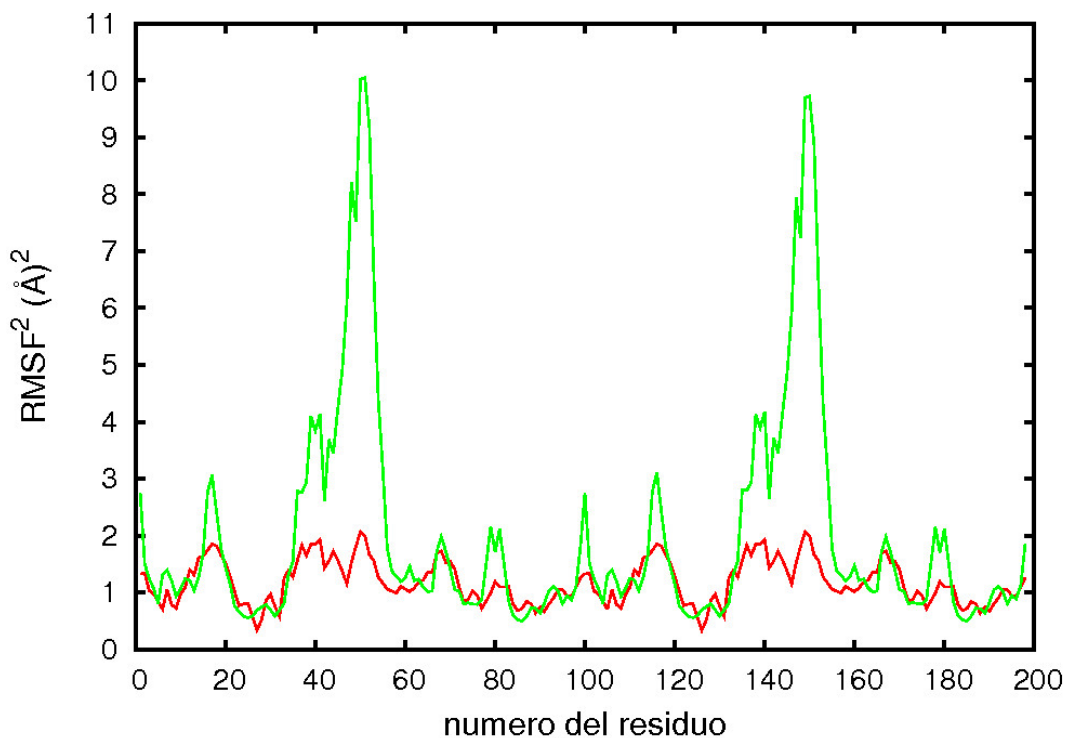


Figura 4.6: Confronto tra il valore di RMSF^2 ottenuto dalla simulazione (linea verde) con le fluttuazioni sperimentali della struttura cristallografica avente 1HHP come codice PDB (linea rossa).

ro). Infatti in questa regione gli RMSF derivati dalla simulazione sono molto maggiori, poiché in una struttura cristallografica i flap non possono avere una grande libertà conformazionale, mentre nella simulazione la libertà conformazionale viene ampiamente esplorata tramite fenomeni di apertura. Discrepanze in altre regioni indicano moti correlati con l'apertura dei flap.

La figura 4.7) riporta l'energia interna (in rosso) e l'energia libera (in verde) in funzione della distanza tra i flap. L'energia interna aumenta con un salto di $\sim 5\text{kcal}$ in corrispondenza di $d \sim 10\text{\AA}$, che può essere considerata la distanza limite che distingue stati aperti e chiusi. La variazione di energia libera di Helmholtz è calcolata tramite

$$\Delta F(d) = -k_B T \ln(N(d)) \quad (4.6)$$

dove $N(d)$ è il numero di configurazioni assunte ad uno specifico valore di d , assumendo un'esplorazione abbastanza sufficiente dello spazio delle fasi.

Come si può osservare, anch'essa sale, ma di meno ($\sim 4\text{ kcal/mole}$) perchè la configurazione con i flap aperti è molto più dinamica di quella con i flap chiusi, ha maggior volume dello spazio configurazionale a disposizione, e quindi il vantaggio entropico $-T\Delta S$ compensa in parte lo svantaggio energetico. È stato mostrato in [58] che alla temperatura $T \sim 350\text{ K}$ il fattore entropico diventa prevalente e la configurazione aperta diventa più stabile di quella chiusa. Tutti i risultati fin qui ottenuti concordano con quelli precedentemente ottenuti nelle referenze [26, 56, 58].

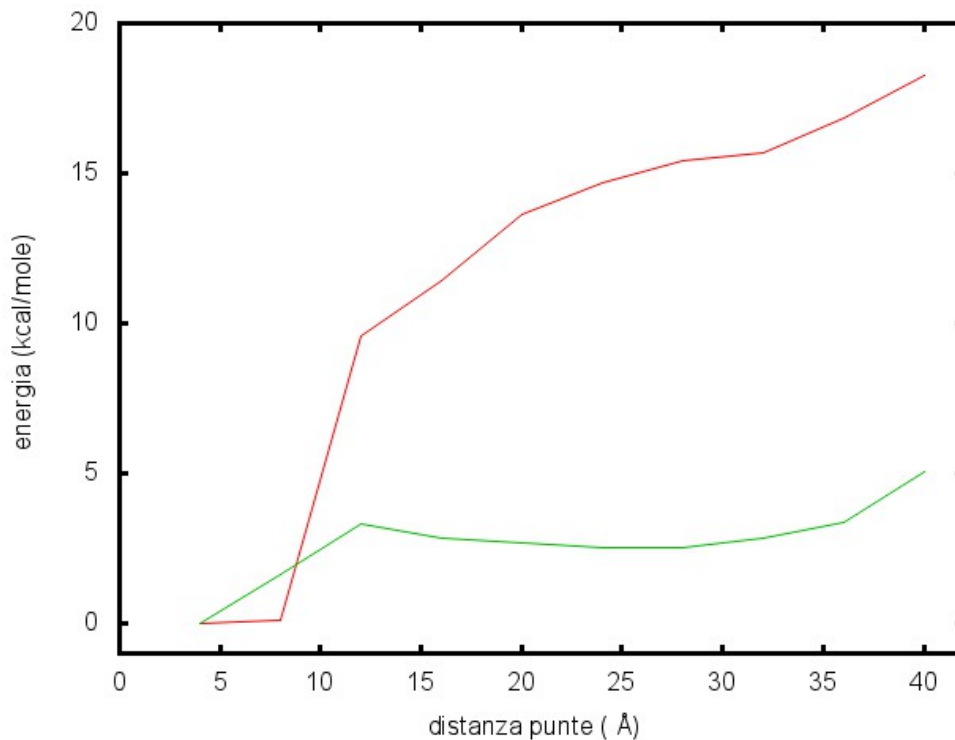


Figura 4.7: Variazione dell'energia in funzione della distanza tra le punte dei flap (Ile50-Ile50'). La linea rossa rappresenta la variazione di energia interna mentre la linea verde l'energia libera. Le due energie sono calcolate mediando su intervalli di 4\AA sulla distanza. La deviazione standard sull'energia interna è valutata in circa $8 - 10$ kcal/mole, mentre sull'energia libera è minore di 0.1 kcal/mole.

4.2.2 Proteasi e substrato

In questa sezione sono riportate simulazioni dell'interazione tra proteasi e un substrato peptidico modello, formato da un nonapeptide di sequenza Lys-Ala-Arg-Val-Leu-Ala-Glu-Ala-Met. L'interazione tra Pr e substrato è descritta come in [26], tramite un termine del tutto simile al termine U_{nb}^{loc} della proteasi isolata: le interazioni sono espresse tramite potenziali morse con parametri strutturali basati sulla struttura sperimentale del complesso PR-substrato, e parametri energetici conformi a quelli del resto della proteina. Tutti i parametri sono riportati nelle appendici D e E che mostra le interazioni non bonded locali tra la proteasi ed il sito di taglio che in questo caso coincide con il substrato.

Simulazione con substrato libero

In una prima simulazione sia la proteasi che il substrato sono stati lasciati entrambi liberi di muoversi. Il sistema è stato simulato utilizzando nuovamente la dinamica di Langevin a 300 K e $\gamma = 2ps^{-1}$, con passo temporale $0.02ps$.

Nella simulazione si osserva che dopo $\sim 1.3 \mu sec$ i flap si aprono per un periodo più lungo (figura 4.9), e questo accade per un'interazione più forte col substrato che poi viene catturato. Successivamente alla cattura, i flap si chiudono più strettamente che non nella proteasi senza substrato (si veda figura 4.3). Inoltre dopo la cattura del peptide si osserva un valore

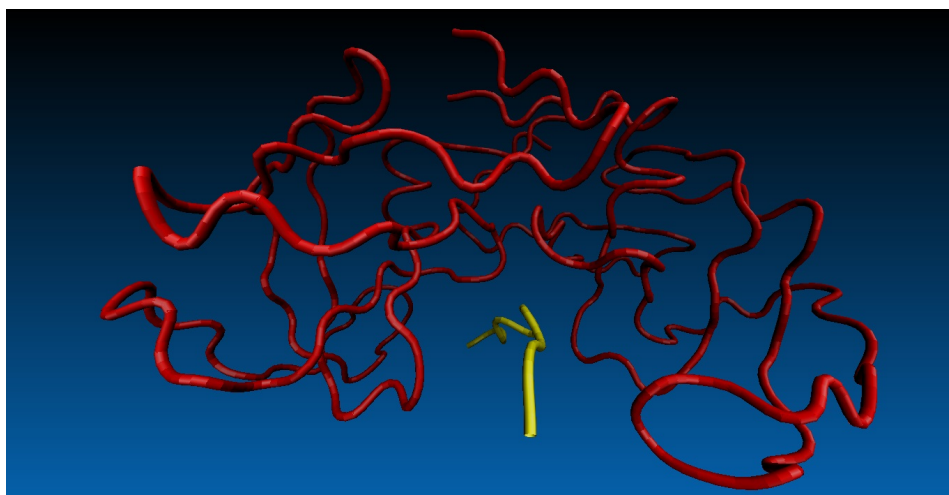


Figura 4.8: Immagine del substrato catturato dalla proteasi.

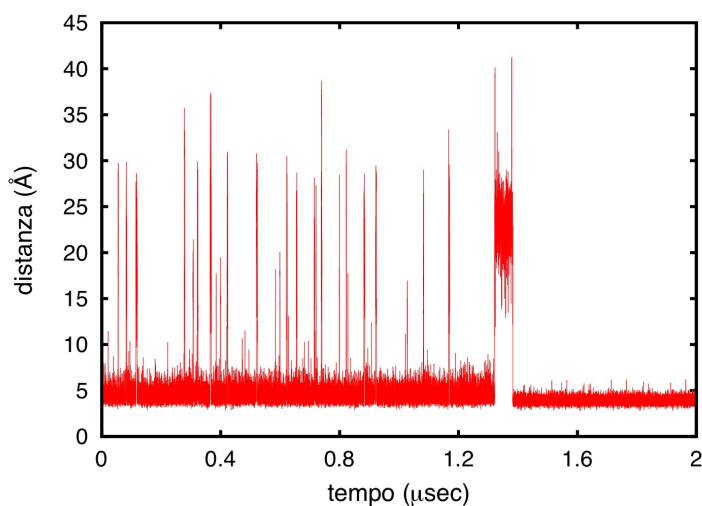


Figura 4.9: Distanza tra le punte dei flap in funzione del tempo della simulazione con il substrato libero. Per distanza tra le punte dei flap si intende la distanza tra gli amminoacidi Ile50 ed Ile50'.

di RMSD più alto rispetto a quello calcolato prima dell'aggancio mostrando ulteriormente la necessità di un cambiamento conformazionale da parte della proteasi (figura 4.10). La cattura del substrato si può osservare anche dai grafici presenti in figura 4.11, dove sono mostrate le variazioni dei vari termini dell'energia a seconda della posizione del substrato.

Notiamo infatti che si ha una diminuzione dell'energia potenziale al momento della cattura del substrato (figura 4.11 b) di ~ 20 kcal/mole. A questo valore contribuiscono sia i termini diretti di legame pr-substrato (figura 4.11 d, linea rossa) che la parte "non bonded" (figura 4.11 d, linea verde) che testimonia una stabilizzazione generale del sistema. I termini

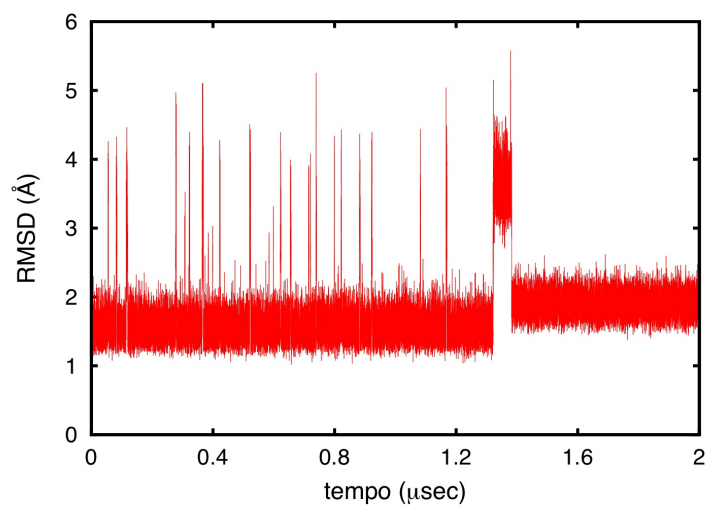


Figura 4.10: RMSD calcolato utilizzando la struttura iniziale come struttura di riferimento nella simulazione con il substrato libero.

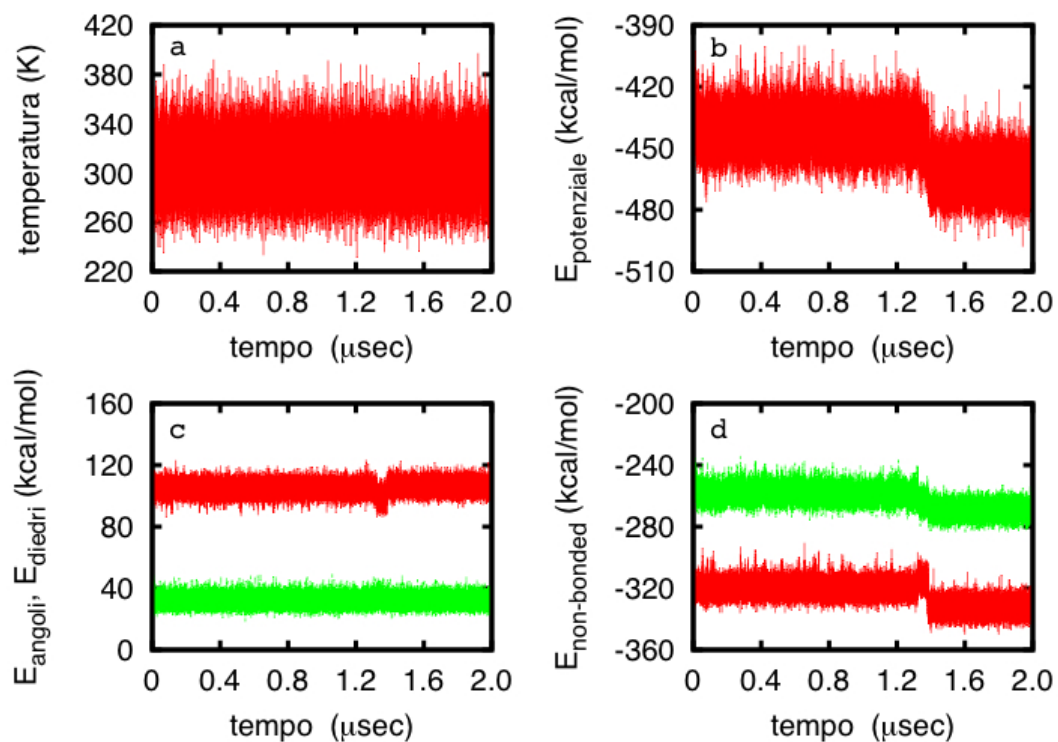


Figura 4.11: Grafici in funzione del tempo della temperatura (a), dell'energia potenziale (b), del potenziale degli angoli (linea rossa in (c)) e dei diedri (linea verde in (c)), e del potenziale dato dai termini non bonded locali (linea rossa in (d)) e dai non bonded non locali (linea verde in (d)) della simulazione con il substrato libero.

conformazionali dell'energia (figura 4.11 c) invece variano poco e strettamente in corrispondenza del momento dell'apertura dei flap che precede il legame. In particolare il termine che descrive gli angoli (linea rossa) appare essere quello che guida la transizione di apertura, mostrando una diminuzione in quel momento. L'energia chimica di legame U_{nb}^{loc} (figura 4.11 d, linea rossa) mostra una leggera barriera di qualche kcal/mole al momento dell'ingresso del substrato nella cavità di taglio, successivamente una diminuzione indicante che il processo di legame è avvenuto correttamente.

La variazione dell'energia interna in funzione della distanza tra la proteasi ed il substrato (figura 4.12), definita come la distanza tra gli amminoacidi Ala6 appartenente al substrato ed Asp25 della triade catalitica dell'enzima, mostra che correttamente all'avvicinarsi del substrato al sito attivo si ha una stabilizzazione.

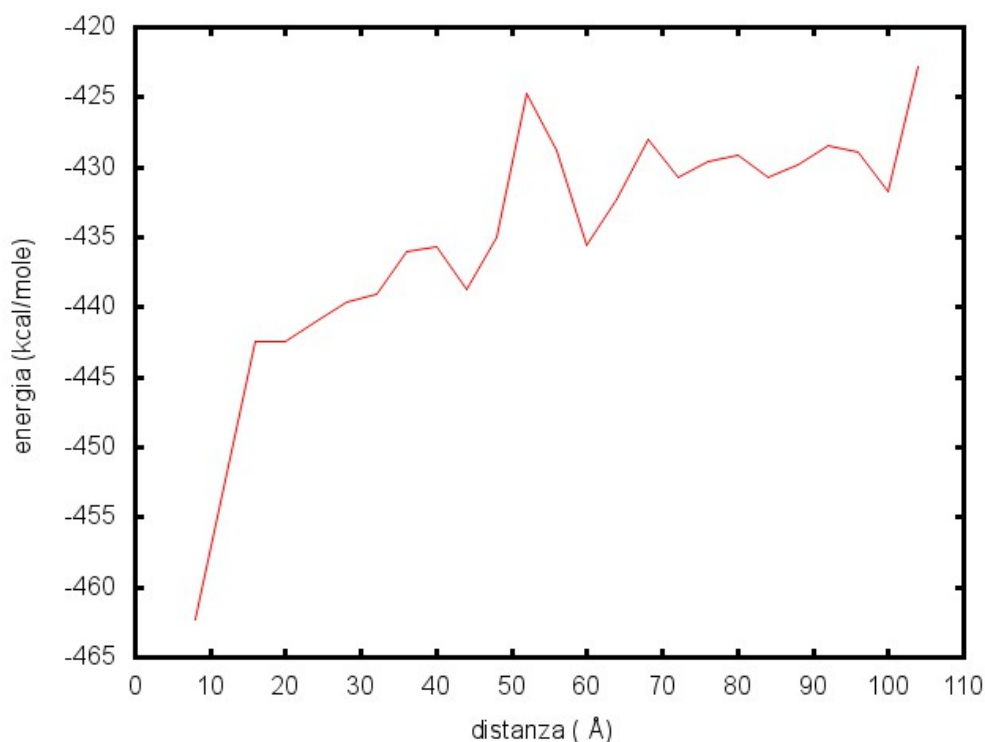


Figura 4.12: Variazione dell'energia potenziale in funzione della distanza tra la proteasi fissata ed il substrato libero in funzione del tempo. Si è presa la distanza tra gli amminoacidi Ala6 appartenente al substrato ed Asp25 della triade catalitica dell'enzima. L'errore statistico su questi dati è di circa 7 – 10 kcal/mole.

Un'altra quantità utile è il raggio di girazione R_g , che rispecchia la dimensione media del complesso e quindi la distanza tra proteasi e substrato. È definito come la radice quadratica media della distanza degli atomi rispetto al centro di massa del complesso [67]:

$$R_g = \sqrt{\frac{\sum_{i=1}^N m_i s_i^2}{\sum_{i=1}^N m_i}} \quad (4.7)$$

Nella simulazione si osserva che il substrato si avvicina alla proteasi molto velocemente, ispeziona lo spazio intorno all'enzima, quando si osserva l'apertura dei flap si ha un lieve

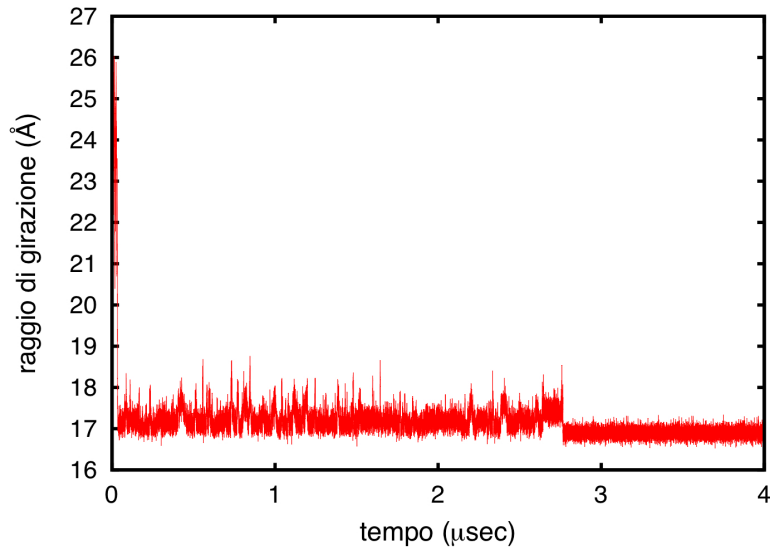


Figura 4.13: Raggio di girazione del complesso substrato-pr nella simulazione con il substrato libero.

aumento di R_g (figura 4.13) e successivamente si ha il minimo valore del raggio di girazione indicante che la cattura è andata a buon fine.

Simulazione con “substrato fisso”

All’interno del virione la Gag non è libera di muoversi, ma è fissata alla parete interna dell’involuppo. Perciò ha un maggior senso fisico una simulazione in cui gli estremi del substrato sono vincolati con un potenziale armonico e la proteasi libera di muoversi. L’uso dei vincoli armonici impedisce il moto diffusivo del substrato, simulando però contemporaneamente la fluttuazione degli estremi. Il valore di γ riguardante la dinamica di Langevin è stato mantenuto uguale a $2ps^{-1}$, così come il passo temporale a $0.02ps$. In questa simulazione si vede l’enzima girare intorno al substrato cercando il giusto punto di legame e infine agganciarsi al substrato.

Dall’analisi dei grafici dell’energia presenti in figura 4.14 si può notare come l’aggancio del substrato avvenga leggermente prima rispetto alla simulazione in cui sia la proteasi che il substrato sono liberi di muoversi, indicando che il vincolo del substrato favorisce il processo di cattura. La stessa cosa si può osservare tramite il grafico che mostra l’andamento della distanza tra le punte dei flap al variare del tempo (figura 4.15) e in quello di RMSD (figura 4.16).

È stata definita la nuova variabile σ per analizzare meglio il processo di aggancio della proteasi al substrato:

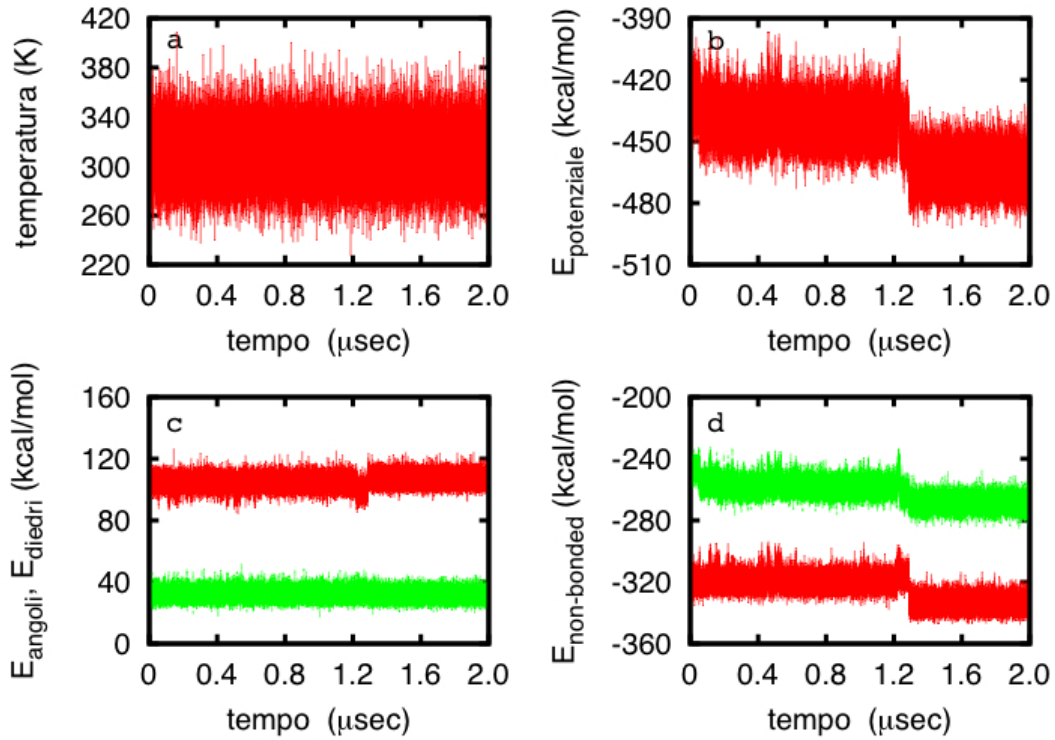


Figura 4.14: Grafici in funzione del tempo della temperatura (a), dell'energia potenziale (b), del potenziale degli angoli (linea rossa in (c)) e dei diedri (linea verde in (c)), e del potenziale dato dai termini non bonded locali (linea rossa in (d)) e dai non bonded non locali (linea verde in (d)) della simulazione con il substrato fisso.

$$\sigma(t) = \sqrt{\frac{1}{N} \sum_{n=1}^N \frac{(r_n(t) - r_n^0)^2}{(r_n^0)^2}} \quad (4.8)$$

In questa equazione N è il numero di tutte le interazioni incluse nel termine U_{nb}^{loc} che descrivono l'interazione chimica tra substrato e Pr: n è in realtà una coppia di indici i, j di cui i appartiene ad una bead della proteasi e j ad una del substrato. $r_n(t)$ è la distanza tra le bead i -esima e j -esima in una data configurazione al tempo t , mentre r_n^0 è il suo valore di riferimento nel complesso Pr-substrato (la lista di tutti i termini inclusi nel calcolo di σ è riportata in appendice E). σ assume il valore 0 quando la configurazione relativa tra proteasi e substrato è quella assunta all'interno del complesso, ovvero quando la proteasi è correttamente agganciata all'interno del complesso. Altrimenti σ assume valori positivi tanto maggiori quanto più la proteasi è lontana dal sito di aggancio. σ monitorizza l'avvicinarsi della proteasi al sito di taglio e se avviene il corretto posizionamento del substrato nel sito attivo dell'enzima, come si vede dalla figura 4.17. La proteasi si avvicina al substrato, esplora i suoi dintorni e successivamente, dopo $\sim 1.2 \mu\text{sec}$ si aggancia. A $\sim 1.3 \mu\text{sec}$ i flap si chiudono posizionando correttamente il substrato e diminuendo ulteriormente σ . Corrispondentemente le punte dei flap passano da una distanza di 11.98 \AA a 4.74 \AA . Un'ulteriore conferma di come la proteasi catturi il substrato la si può ottenere dal raggio di girazione del complesso pr-substrato (fi-

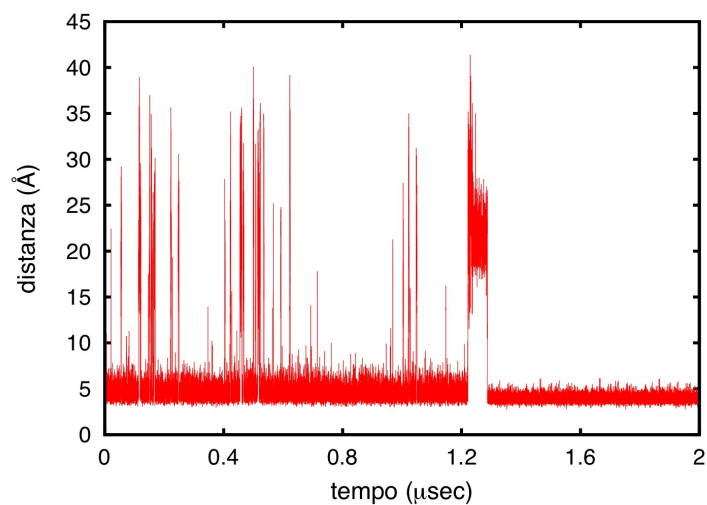


Figura 4.15: Distanza tra le punte dei flap in funzione del tempo della simulazione con il substrato fisso. Per distanza tra le punte dei flap si intende la distanza tra gli amminoacidi Ile50 ed Ile50'.

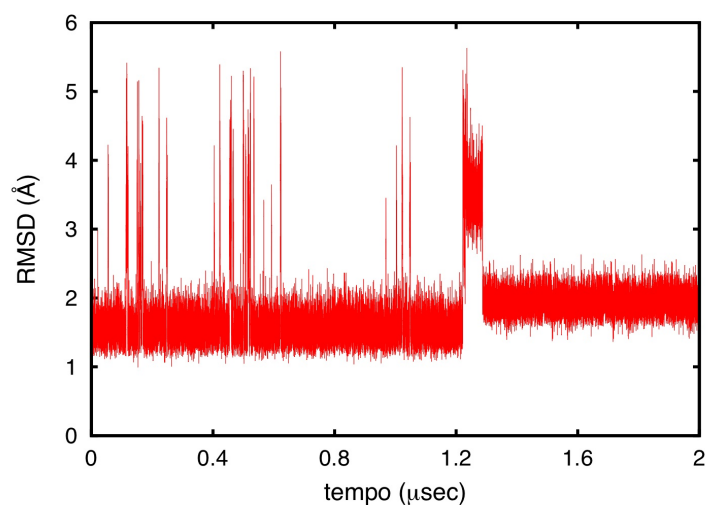


Figura 4.16: RMSD calcolato utilizzando la struttura iniziale come struttura di riferimento nella simulazione con il substrato fisso.

gura 4.18): la sua diminuzione misura la riduzione della dimensione del complesso e quindi la sua corretta formazione.

Un fatto interessante che si vede grazie al calcolo di $RMSF^2$, è l'ulteriore aumento delle oscillazioni nelle regioni dei flap indotto dall'interazione tra la proteasi e il substrato (figura

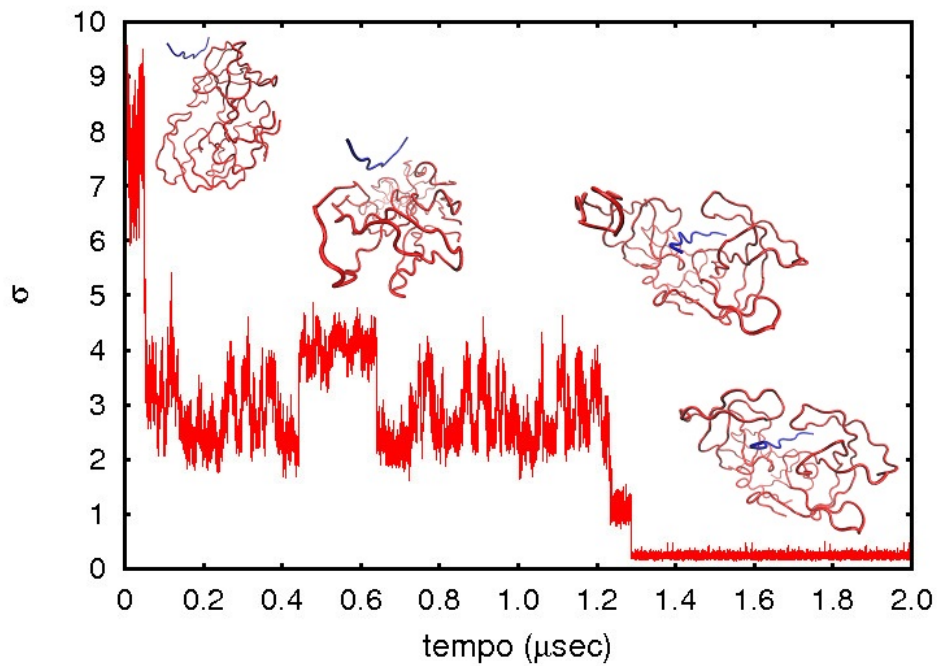


Figura 4.17: Grafico della nuova variabile σ in funzione del tempo della simulazione della proteasi libera con il substrato fissato alle due estremità.

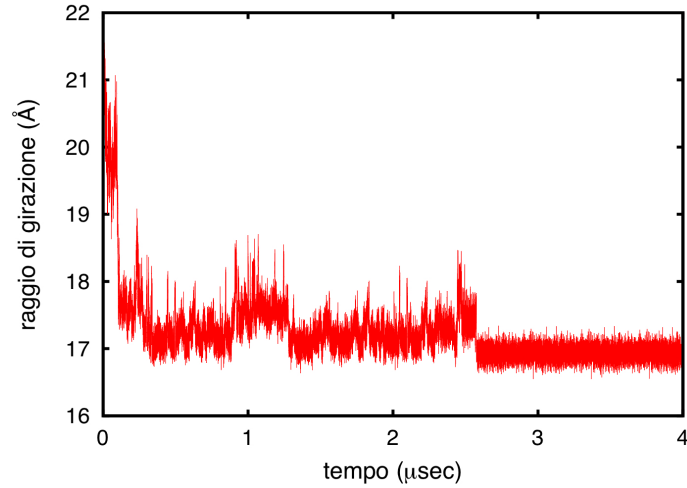


Figura 4.18: Raggio di girazione del complesso substrato-pr nella simulazione con il substrato libero.

4.19). Inoltre viene a crearsi un'asimmetria tra le due regioni dei flap, che è in accordo con un'ipotesi suggerita da precedenti studi teorici secondo cui l'interazione tra pr e substrato è asimmetrica sia staticamente che dinamicamente [24, 26].

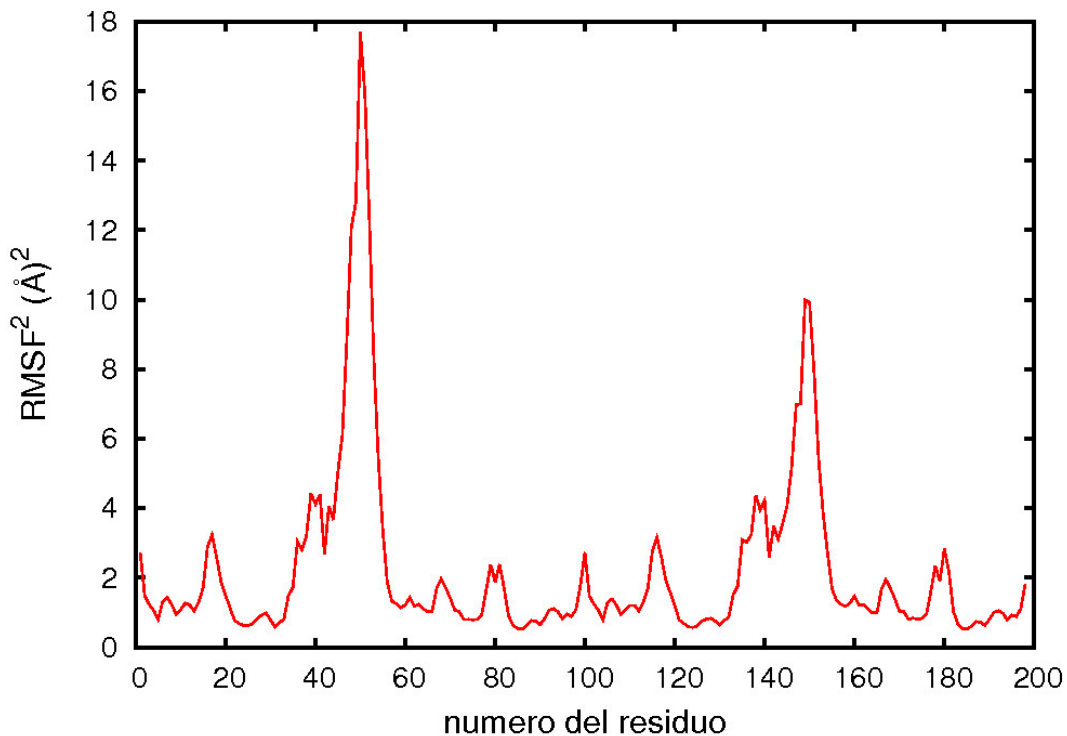


Figura 4.19: RMSF^2 calcolato dalla simulazione dove la proteasi si lega con il substrato fissato agli estremi.

4.3 Ottimizzazione del campo di forze di Gag

In questa sezione verrà descritta la costruzione del modello della poliproteina Gag del virus HIV-1, basato su dati ottenuti da un lavoro di risonanza magnetica sulla struttura MA-CA_{NTD} contenente due domini consecutivi di Gag. Verrà discussa la scelta sulla struttura di riferimento di partenza per le simulazioni di dinamica molecolare. Verrà illustrata la costruzione del campo di forze includente flessibilità sufficiente nel collegamento tra i domini MA e CA_{NTD} . Per ottenere questa flessibilità in accordo con i dati sperimentali a nostra disposizione delle strutture NMR, è stato così introdotto un modello bistabile nella regione di collegamento tra MA e CA_{NTD} .

4.3.1 Dati NMR

Per prima cosa si è selezionato lo specifico sito di taglio da studiare. Non essendoci strutture cristallografiche dell'intera poliproteina Gag, si sono cercate strutture parziali che contenessero almeno due domini adiacenti ed un sito di taglio tra essi. La scelta è ricaduta sulla sequenza MA-CA_{NTD} , di cui è stato trovato un lavoro che ne analizzava la struttura per mezzo della risonanza magnetica (NMR) [64]. In questo lavoro è stata infatti determinata la struttura di un frammento della poliproteina Gag di HIV-1, chiamata Gag^{283} che include i domini MA e CA_{NTD} , prima del processo di maturazione (in totale 288 amminoaci-

di). La determinazione sperimentale (codice PDB: 1L6N), contiene 20 diverse strutture che testimoniano la flessibilità strutturale della Gag.

4.3.2 Scelta strutture

Dal momento che nelle 20 strutture l'orientamento dei domini è variabile, è stata fatta una scrematura per scegliere una configurazione di partenza per le simulazioni. Per prima cosa sono state scartate le strutture in cui i domini MA e CA_{NTD} erano vicini tra di loro. Infatti nel virione immaturo di HIV-1, i vari domini di Gag formano una struttura a strati in cui le Gag si estendono dalla membrana virale fino al centro del virione in maniera radiale (figura 4.20). Infatti l'estremità N-terminale di MA è legata allo strato interno della membrana [12]. Sono state così eliminate le strutture che non supportavano una tale disposizione dei domini, in favore di quelle più estese in cui i domini MA e CA erano lontani.

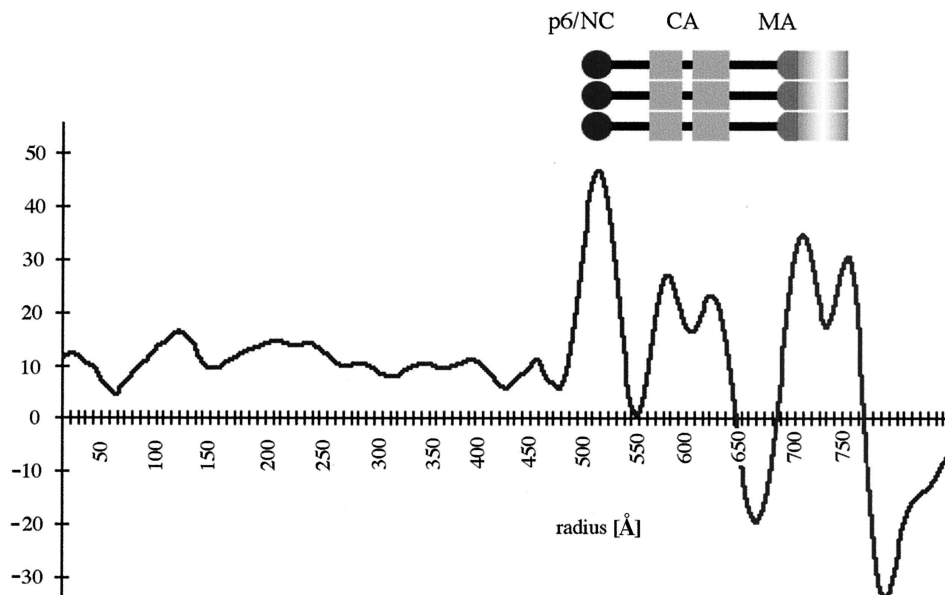


Figura 4.20: Media di 10 distribuzioni di densità della poliproteina Gag di HIV-1 di tipo naturale che mostra la presenza di un arrangiamento radiale dei domini della Gag. Sono presenti dei picchi di densità corrispondenti ai vari domini della poliproteina. Il grafico mostra la densità (asse y) in funzione della distanza dal centro della particella virale in angstrom (asse x) [11].

Più precisamente da studi di cryoEM (criomicroscopia elettronica) si è visto [11, 12, 8] che il dominio CA si trova ad una distanza di 100\AA dalla membrana, in particolare la regione C-terminale di MA e la regione N-terminale di CA formano una struttura estesa di circa 70\AA di lunghezza. Inoltre è stato mostrato che la regione C-terminale del dominio MA dopo la maturazione forma un'elica- α comprendente 25 amminoacidi. Questa elica parte dall'inizio del dominio MA e dovrebbe estendersi verso il centro del virione. La sua lunghezza di 37\AA dovrebbe occupare la maggior parte dei 40\AA di spazio osservato tra i domini MA e CA_{NTD} visti tramite microscopia elettronica nei virioni immaturi di HIV-1.

Tramite la conoscenza della distanza tra i due domini MA e CA_{NTD} , è stata calcolata la distanza tra i centri di massa dei due domini relativa alle varie strutture, per eliminare i domini situati troppo vicini o troppo lontani.

Per quanto riguarda MA, il centro di massa è stato calcolato dal primo amminoacido (*Gly1*) fino a *Thr121* (figura 4.21). Invece per CA_{NTD} il centro di massa è stato calcolato da *Ile146* a *Ser280*. Sono state tenute le strutture che avevano una distanza tra i centri di massa dei due domini compresa in un range tra $70 - 90\text{\AA}$ (figura 4.21).

I domini MA e CA_{NTD} sono collegati tramite un linker flessibile nel virione immaturo. Questo linker contiene il sito di taglio della Gag, composto dalla sequenza di otto amminoacidi dalla posizione 128 alla posizione 135: Ser-Gln-Asn-Tyr-Pro-Ile-Val-Gln. Questo ha portato all'eliminazione finale delle strutture che avevano una parte del linker troppo contratta, che risulta non fisica considerando la struttura che la Gag assume nel virione, e avrebbe potuto impedire l'aggancio della proteasi. Sono così rimaste due strutture candidate come riferimento per le simulazioni, che sono indicate con i numeri dei corrispondenti modelli all'interno della determinazione sperimentale NMR (2 e 13) e riportate in figura 4.21.

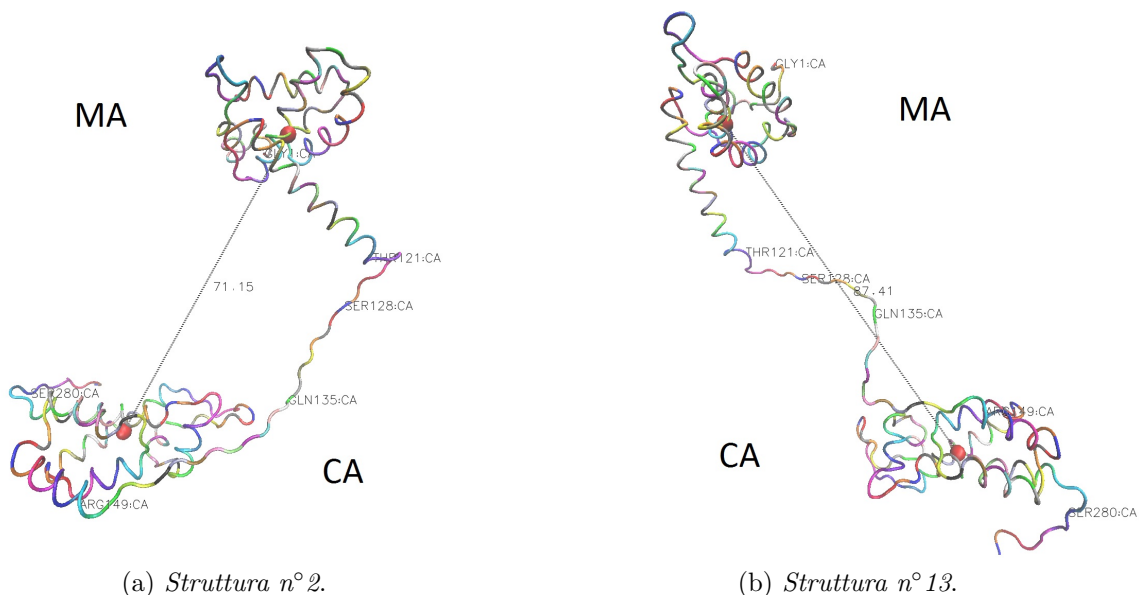


Figura 4.21: Struttura MA- CA_{NTD} . MA è compresa da Gly1 a Thr121. CA_{NTD} è compresa da Arg149 a Ser280. Le palline rosse rappresentano i centri di massa del dominio, e la loro distanza (linea tratteggiata) è misurata in angstrom. La regione tra le due proteine è il linker, dove è possibile osservare il sito di taglio: da Ser128 a Gln135.

Si è poi valutato in quali regioni della proteina ci fossero più variazioni strutturali, tramite il calcolo dell'RSMD relativo alle 20 strutture prese dal codice pdb 1L6N, ognuna delle quali è stata divisa in tre regioni:

1. da Gly1 a Thr121 (dominio MA)
2. da Gly122 a Pro148 (linker)
3. da Arg149 a Ser280 (dominio CA)

L'RMSD di ciascuna delle tre regioni è stato valutato separatamente. I valori medi ottenuti sono 2.057, 8.138, 1.794 per MA, linker e CA rispettivamente (si veda appendice F per i dettagli). Il linker risulta avere RMSD maggiore, ed infatti è la regione maggiormente destrutturata. Per questo motivo la sua modellizzazione necessita una particolare attenzione.

4.3.3 Costruzione del campo di forze di Gag

Per una prima versione del campo di forze della poliproteina Gag, è stato usato un modello minimalista con qualche modifica rispetto a quello usato per la proteasi, con cambiamenti che sono presenti in particolar modo nel potenziale dato dagli angoli e in quello dato dai diedri. I valori dei parametri sono stati usati in un precedente modello simile applicato ad altre proteine [68, 69] e sono descritti di seguito.

Il potenziale dato dagli angoli, U_θ , ha la funzione di un coseno armonico:

$$U_\theta = \frac{k}{2} (\cos(\theta) - \cos(\theta_0))^2$$

con il valore di θ_0 che dipende dalla struttura secondaria usata per la costruzione del campo di forze. Per quanto riguarda i vari k , essi dipendono dal valore di θ secondo la relazione

$$k = 3000 \left(\frac{\sin(\theta/107)}{\theta/107} \right)^2 + 10$$

Il potenziale degli angoli diedri U_ϕ è invece rappresentato dalla funzione coseno:

$$U_\phi = A[1 + \cos(\phi - \delta)]$$

dove δ è ricavato dalle strutture conosciute della Gag e A può assumere due valori, $A = 20$ se $|\phi| > 100$, $A = 3$ altrimenti.

Per le interazioni non-bonded locali è stato usato il potenziale di tipo Morse come quello dell'eq. 4.2 dove però i parametri k ed E_0 avevano una diversa dipendenza da r rispetto al modello della proteasi:

$$E_0 = 3.8 \exp\left(-\left(\frac{r}{6}\right)^8\right) + 0.05$$

$$k = 0.7 + 0.7 \exp\left(-\left(\frac{r}{6}\right)^8\right)$$

Le interazioni locali e non locali sono separate tramite un cutoff di 8.5 Å. Anche per le interazioni non-bonded non locali è stato utilizzato un potenziale di tipo morse con la distanza di equilibrio r_0 settata a 9.5 Å, il range di interazione $k = 0.7 \text{ Å}^{-1}$ e la forza $E_0 = 0.05 \text{ kcal/mole}$.

Per simulare i vincoli realmente presenti sui domini MA e CA_{NTD} all'interno del virione (ancoraggio di MA all'involuppo e legame di CA_{NTD} al resto della Gag), è stato fissato l'amminoacido di CA_{NTD} più vicino al baricentro, Lys262, e sono stati fissati tre atomi di MA (Leu20, Trp35 e Glu72). I vincoli alle posizioni iniziali sono mantenuti tramite potenziali armonici con costante $k = 50 \text{ kcal/mole}$.

4.3.4 Dinamica di Gag da sola e miglioramento del campo di forze

Sono state così eseguite due simulazioni di Gag, una avente come configurazione di partenza la struttura numero 2 e l'altra la struttura numero 13, per analizzarne il comportamento. Il campo di forze è stato inizialmente ricavato dalla struttura numero 18, la quale assumeva un'ideale configurazione estesa. Come si vede dall'analisi degli angoli nella figure 4.22, i valori delle strutture trovate dalle due simulazioni sono abbastanza simili tra loro, ma differiscono notevolmente nella regione del linker (da Gly122 a Pro148) dalla distribuzione delle 20 strutture NMR (in blu). I risultati riguardo la varianza degli angoli sono simili a quelli ottenuti per la media degli angoli e sono riportati in appendice H.

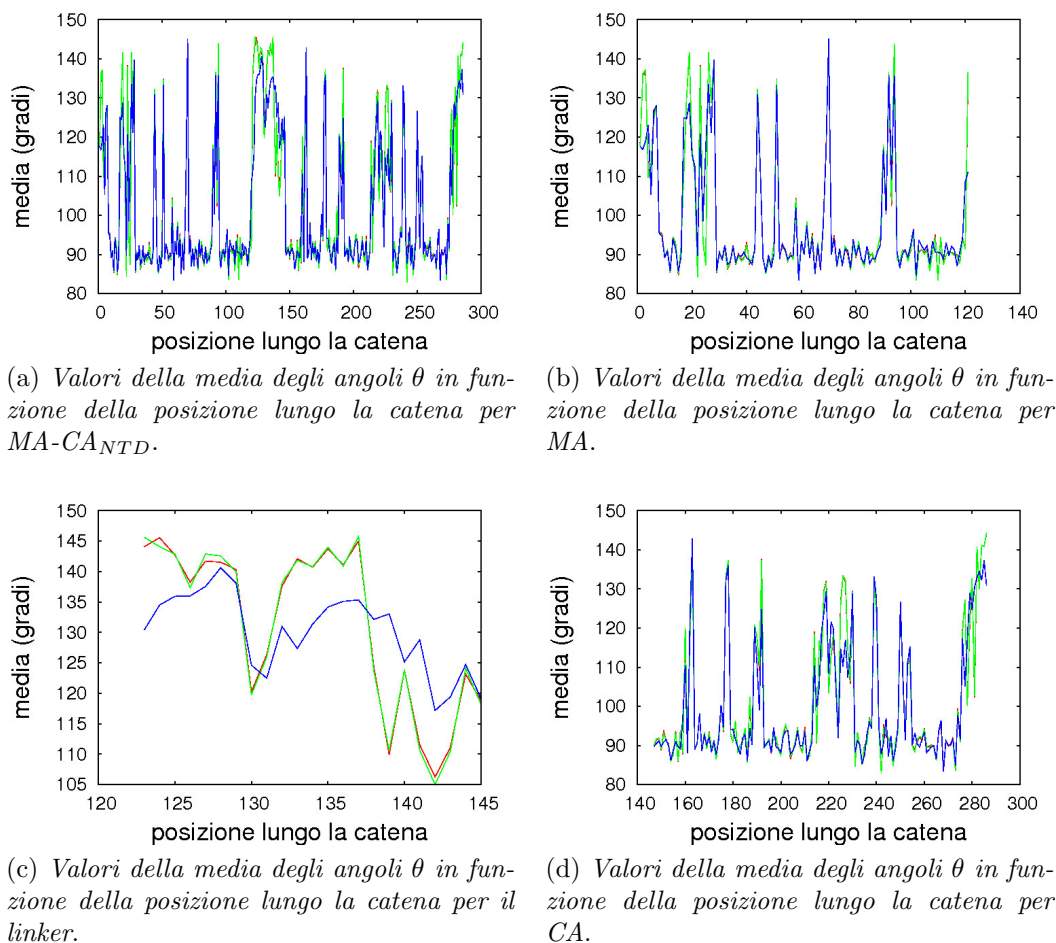


Figura 4.22: Confronto tra la simulazione avente la struttura 2 come configurazione di partenza(rosso), quella avente la struttura 13(verde) e le strutture NMR 1L6N(blu).

Dall'analisi degli angoli e dei diedri delle 20 strutture ricavate dalla risonanza magnetica presenti nella seconda regione (da Gly122 a Pro148)(vedi appendice G), è stato osservato che in questa regione poteva essere utilizzato un modello a doppia buca sia per gli angoli che per i diedri. In questo modo si cercava di ottenere dalle simulazioni un insieme di strutture che rappresentassero più accuratamente i dati sperimentali.

Sono stati usati due differenti modelli a doppia buca per gli angoli, modelli che sono stati messi a confronto per stabilire quale dei due fosse il più corretto da utilizzare. Il

primo modello si basa su un potenziale quartico (eq 4.1), mentre il secondo modello è una combinazione di due potenziali.

Date due funzioni u_1 e u_2 che descrivono ognuna una buca di potenziale, la doppia buca di potenziale è ottenuta dalla loro combinazione:

$$u_{dw} = \frac{u_1 + u_2 - \Delta}{2} - \frac{1}{2} \sqrt{(u_1 - u_2 + \Delta)^2 + \varepsilon^2} \quad (4.9)$$

dove Δ è la stabilità relativa tra le due buche di potenziale ed ε serve a smussare la cuspidè che si forma nella zona centrale (i dettagli della funzione sono riportati in figura 4.23).

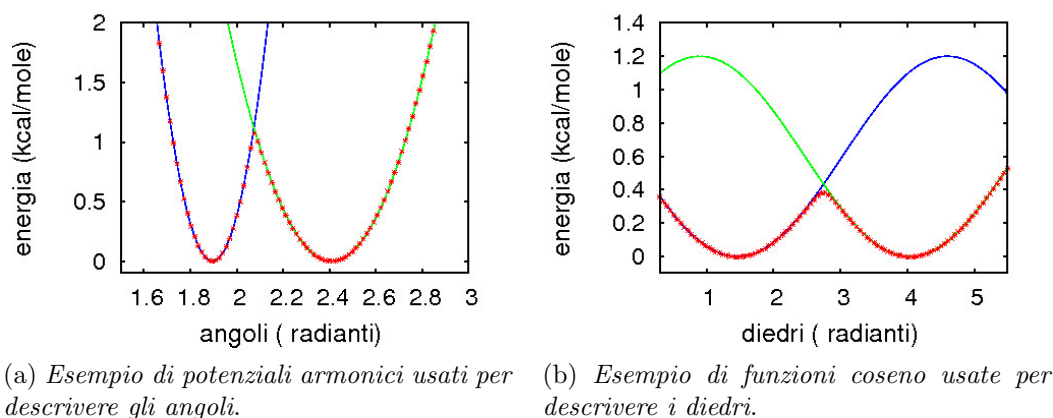


Figura 4.23: Grafico del potenziale a doppia buca, in rosso tratteggiato, composto da due potenziali di partenza, in verde ed in blu.

Dalle simulazioni è stato constatato tramite l'analisi delle distribuzioni angolari, di cui si ha un esempio in figura 4.24, che il secondo metodo (linea rossa) rappresenta in maniera migliore i dati sperimentali (linea blu) rispetto al metodo con il potenziale quartico (linea verde). Inoltre è applicabile a tutti i termini del potenziale, in particolare a quello diedrale. Nelle appendici I e J sono riportate le varie distribuzioni degli angoli e dei diedri della regione del linker.

Per costruire il potenziale a doppia buca degli angoli, le funzioni u_1 e u_2 erano dei potenziali armonici, mentre per la doppia buca relativa ai diedri sono state prese delle funzioni coseno. I minimi delle doppie buche sono stati ricavati utilizzando le strutture NMR: fissato un valore limite per ogni posizione degli angoli (o dei diedri) lungo la catena, in modo da selezionare a quale delle due buche appartenessero gli angoli (o i diedri) delle varie strutture, veniva calcolata la media dei vari angoli (o diedri) per ottenere il valore minimo delle buche. Inizialmente l'altezza relativa tra le due buche Δ è stata posta uguale a 0 sia per gli angoli che per i diedri e successivamente è stata modificata per gli angoli per bilanciare meglio il rapporto delle strutture ottenute dalle simulazioni sempre in accordo con i dati sperimentali. L'altezza della barriera della doppia buca è stata posta di 0.2 kcal/mole per i diedri mentre variava per gli angoli e il parametro ε è stato posto a 0.1 kcal/mole. I valori finali delle distribuzioni degli angoli e dei diedri del linker sono quelli ottenuti dal modello bistabile riportati nelle appendici I e J. Il modello riproduce abbastanza bene le distribuzioni angolari e diedrali del linker, considerata la difficoltà del confronto dovuta dalla bassa statistica dei dati NMR.

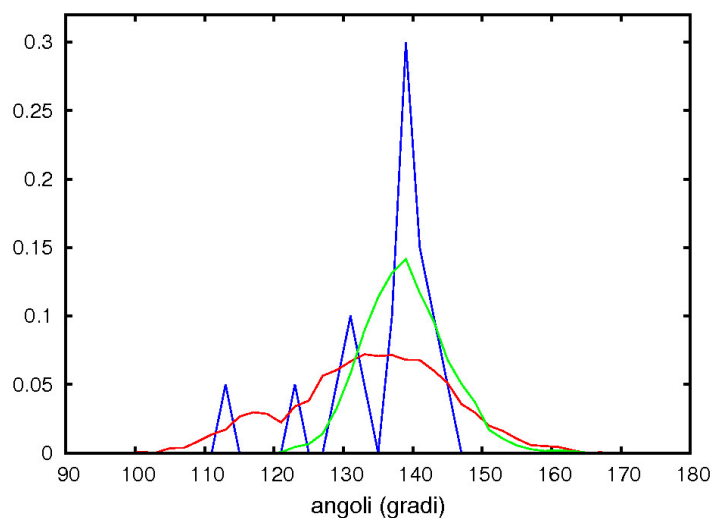


Figura 4.24: Distribuzioni normalizzate degli angoli aventi posizione 126 nella catena. La linea rossa rappresenta i valori ottenuti dalla simulazione usante il modello bistabile implementato sia sugli angoli che sui diedri del linker, la linea verde i dati ricavati dalla simulazione con il potenziale quartico applicato agli angoli del linker, e la linea blu i dati delle strutture NMR.

4.4 Simulazioni della proteasi con Gag

4.4.1 Campo di forze dell'interazione Gag-Proteasi

Il campo di forze che descrive l'interazione Gag-Proteasi è basato su quello delle precedenti simulazioni tra proteasi e substrato modello, ora con delle varianti. Le interazioni specifiche tra proteasi e sito di taglio della Gag sono descritte con potenziali di tipo Morse, i cui parametri sono stati prima scelti come in [26] e poi variati per migliorare la dinamica di interazione (seconda colonna della tabella 4.1). In particolare tenendo fissi i valori delle distanze di equilibrio, precedentemente ottimizzati, sono state variate le energie di legame dei potenziali Morse, in modo da dare pesi relativi diversi alle interazioni specifiche (tra la proteasi e il sito di taglio della Gag) e non specifiche tra Pr e Gag. I campi considerati sono riportati in tabella 4.1 e graficati in figura 4.25.

4.4.2 Simulazioni dell'interazione Pr-Gag

I quattro campi di forze indicati nella sezione precedente sono stati utilizzati in diverse simulazioni (figura 4.26) a partire da diverse strutture iniziali:

1. Struttura del complesso formato, con la proteasi agganciata al sito di taglio.
2. Struttura con Gag e proteasi separate.
3. Struttura con proteasi agganciata alla Gag, ma in posizione spostata rispetto al sito di taglio.

Le strutture iniziali per 1) sono state ottenute allineando il substrato modello (il nonapeptide) del complesso cristallografico al corrispondente sito di taglio della Gag e riottimizzando

n° campo	interazioni i-j tra PR-Gag specifiche	interazioni PR-Gag non specifiche
1° FIELD	$E_0 = 6 \exp(-r_0/2.822)$ $k = 0.70711$	$E_0 = 0.20708$ $k = 0.70711$
2° FIELD	$E_0 = 6 \exp(-r_0/2.822)$ $k = 0.70711$	$E_0 = 0.10354$ $k = 0.70711$
3° FIELD	$E_0 = 4 \exp(-r_0/2.822)$ $k = 0.70711$	$E_0 = 0.10354$ $k = 0.70711$
4° FIELD	$E_0 = 3 \exp(-r_0/2.822)$ $k = 0.70711$	$E_0 = 0.10354$ $k = 0.70711$

Tabella 4.1: Lista dei FIELD usati nelle simulazioni Gag-PR.

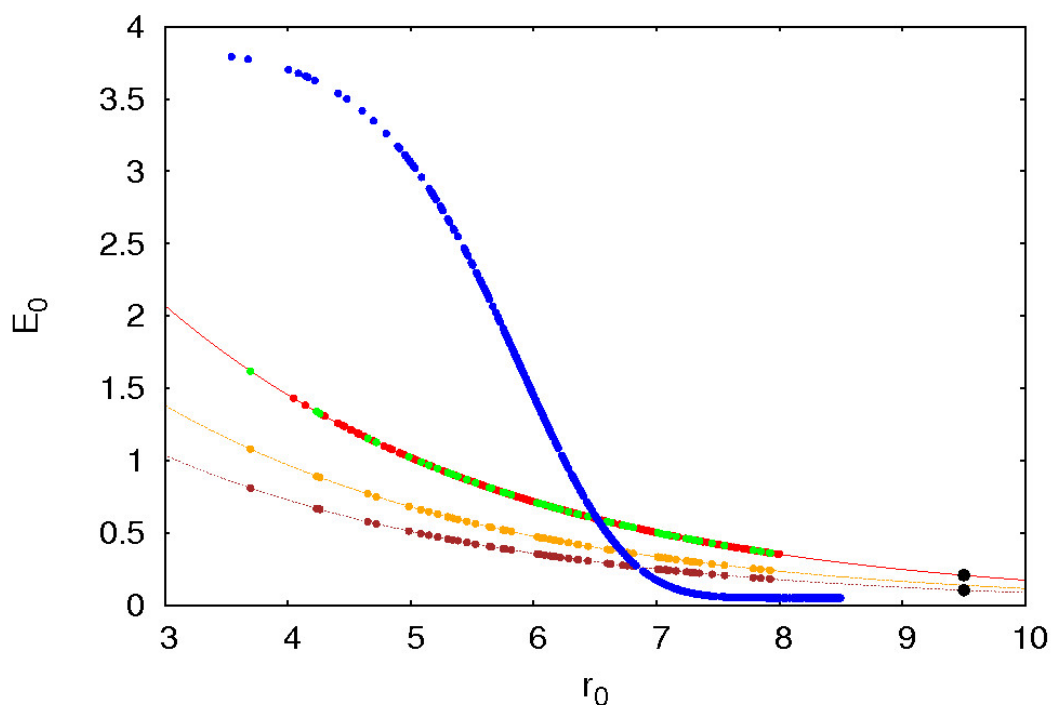


Figura 4.25: Valori dei parametri E_0 in funzione dei parametri r_0 dei legami non-bonded. In rosso sono rappresentati i punti corrispondenti alle interazioni locali PR-PR ed in blu Gag-Gag. Le interazioni locali PR-Gag hanno colori diversi in base al campo di forze utilizzato, dove i punti mostrano le interazioni specifiche e le curve i rispettivi andamenti esponenziali: per il FIELD n°1 e n°2 verde, per il FIELD n°3 arancione ed per il FIELD n°4 marrone. La curva rossa mostra che le interazioni locali PR-PR e PR-Gag dei FIELD n°1 e n°2 seguono lo stesso andamento esponenziale. I pallini neri in basso a destra rappresentano i valori scelti per i legami van der Waals.

localmente, cioè sono state scelte due strutture con il valore di $RMSD < 0.8$ calcolato lungo il sito di taglio (chiamate '2457' e '3002'). Le strutture per 2) e 3) sono state ricavate dalla 1) semplicemente spostando la proteasi, in modo che il suo sito attivo non contenesse il linker in 2) e lungo il linker nel caso 3). Tutte le simulazioni sono state fatte con lo stesso protocollo,

n° campo	pr attaccata al sito di taglio		pr non attaccata al sito di taglio	pr spostata lungo il linker		
	struttura 2457	struttura 3002	struttura 2457	2 aa	4 aa	4 aa *
1° FIELD	pr rimane attaccata al sito di taglio	pr rimane attaccata al sito di taglio	flap impigliato con MA			
2° FIELD	pr rimane attaccata al sito di taglio	pr rimane attaccata al sito di taglio	la pr è attratta da CA			
3° FIELD	pr rimane attaccata al sito di taglio	pr rimane attaccata al sito di taglio	pr si attacca al sito di taglio	pr scivola verso il sito di taglio	pr scivola verso il sito di taglio	pr si stacca da Gag e si riattacca successivamente al sito di taglio
4° FIELD	pr rimane attaccata al sito di taglio	pr rimane attaccata al sito di taglio	pr non si aggancia			

Figura 4.26: Nella tabella sono riportate le varie simulazioni eseguite per verificare l'interazione tra proteasi e Gag. La struttura di partenza di Gag utilizzata per le simulazioni con la pr spostata lungo il linker, è stata ricavata dalla struttura 2457 durante la simulazione in cui pr si attacca al sito di taglio. 2aa e 4aa indicano di quanti amminoacidi la pr è spostata dal sito di taglio, mentre 4aa* è stata ricavata da 4aa dopo un riassetamento dell'intero sistema Gag-pr.

$\gamma = 2ps^{-1}$ e passo temporale $0.02ps$, ma hanno lunghezze variabili a seconda della necessità.

Il caso 1) è stato studiato con tutti campi di forze, i quali danno tutti esito positivo: la struttura del complesso viene ben riprodotta e il complesso rimane stabile per tempi relativamente lunghi. La figura 4.27 riporta alcune strutture estratte a campione dalla simulazione con il campo Field n°3. Tutte le altre sono riportate in appendice K, che contiene anche il monitoraggio delle energie e temperature lungo i run, gli rsmf e la variabile σ calcolati in queste simulazioni.

Più interessanti sono le simulazioni del caso 2) come si vede dalla tabella in figura 4.26, con l'aggancio della proteasi che non avviene in tutti i casi, ma solo quando si ha un preciso rapporto tra le energie di interazione pr-gag specifiche del sito di taglio e le altre, e quando le prime hanno un peso maggiore delle seconde. Questo peso maggiore simula il riconoscimento sequenziale da parte della proteasi rispetto al sito di taglio, e quindi ha un significato fisico. Questo avviene in particolare usando il FIELD n°3. Di seguito è analizzata la dinamica tratta dalla simulazione con questo campo di forze.

In figura 4.28 sono riportati i vari andamenti dei termini delle energie in funzione del tempo, i quali mostrano una netta diminuzione dell'energia potenziale e dell'energia dei termini non-bonded locali dopo l'aggancio dell'enzima con il sito di taglio, e una momen-

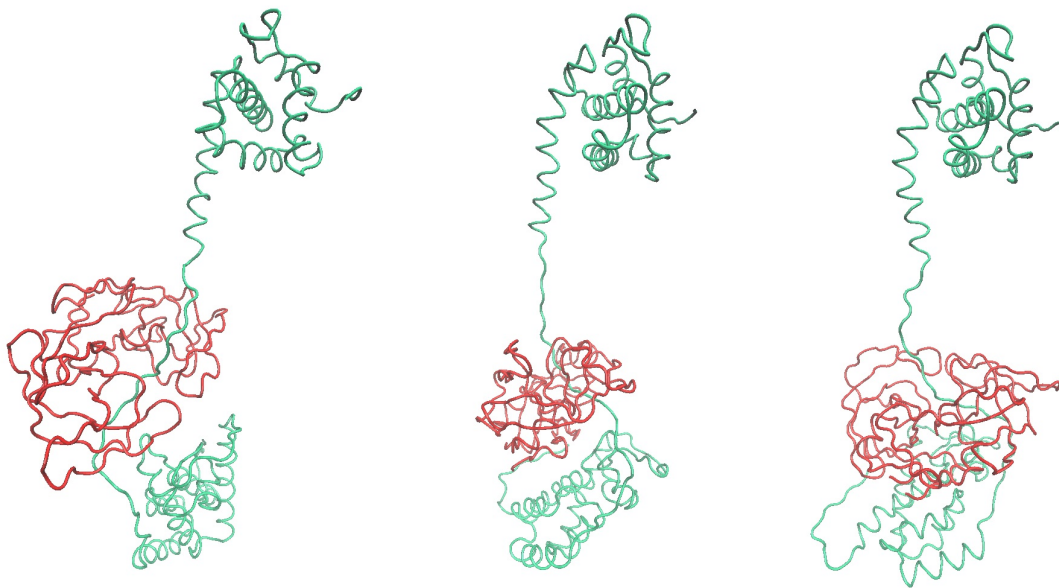


Figura 4.27: Strutture prese dalla simulazione della proteasi agganciata al sito di taglio con il FIELD n°3 e struttura di partenza di Gag '2457'.

tanea diminuzione dell'energia data dagli angoli nel momento dell'aggancio, come è stato precedentemente visto nelle interazioni tra la proteasi ed il substrato modello.

Dai grafici riguardanti l'apertura dei flap rispetto al tempo e di RMSD (figure 4.29 a) e c)), si possono osservare differenti ampi fenomeni di apertura dei flap della proteasi, presumibilmente dovuti al maggior spazio conformazionale che l'enzima deve analizzare prima di riuscire ad agganciarsi al sito di taglio ($\sim 0.85\mu s$) rispetto a quando la proteasi si lega al substrato modello.

È stata riutilizzata la nuova variabile σ definita precedentemente nell'eq.4.8, utilizzando le interazioni non-bonded tra la PR e il sito di taglio della struttura MA-CA_{NTD} (vedi appendice E), dove gli amminoacidi del complesso MA-CA_{NTD} coinvolti nelle interazioni specifiche tra la proteasi e la Gag, sono quelli compresi dalla posizione 127 alla posizione 135. Si può osservare in figura 4.30 come la proteasi ispezioni il linker cercando il punto favorevole all'aggancio con la Gag, il quale avviene a $\sigma \sim 2$ seguito da una netta diminuzione di σ che scende quasi a 0 quando avviene la chiusura dei flap attorno al sito di taglio. Il picco $\sigma \sim 16$ è dovuto al momentaneo allontanamento della proteasi dalla Gag. La variazione della distanza tra l'enzima e la poliproteina è anche mostrato in figura 4.29 d), che riproduce il raggio di girazione del complesso Gag-Pr al variare del tempo.

Una delle conseguenze date dagli ampi fenomeni di apertura dei flap è l'aumento delle oscillazioni nelle regioni dei flap rispetto a quanto visto nell'interazione tra la proteasi ed il substrato. Infatti il valore di RMSF² risulta maggiore nell'interazione tra Gag e proteasi (figura 4.29 b)) rispetto ai valori di RMSF² visti fino ad ora.

Se invece il peso delle interazioni Gag-pr non specifiche è troppo alto rispetto a quelle specifiche, come nel caso del campo di forze n°4, non si vede l'aggancio. Se il peso relativo delle specifiche è alto ma sono entrambe forti, come nei campi n°1 e n°2, la proteasi tende

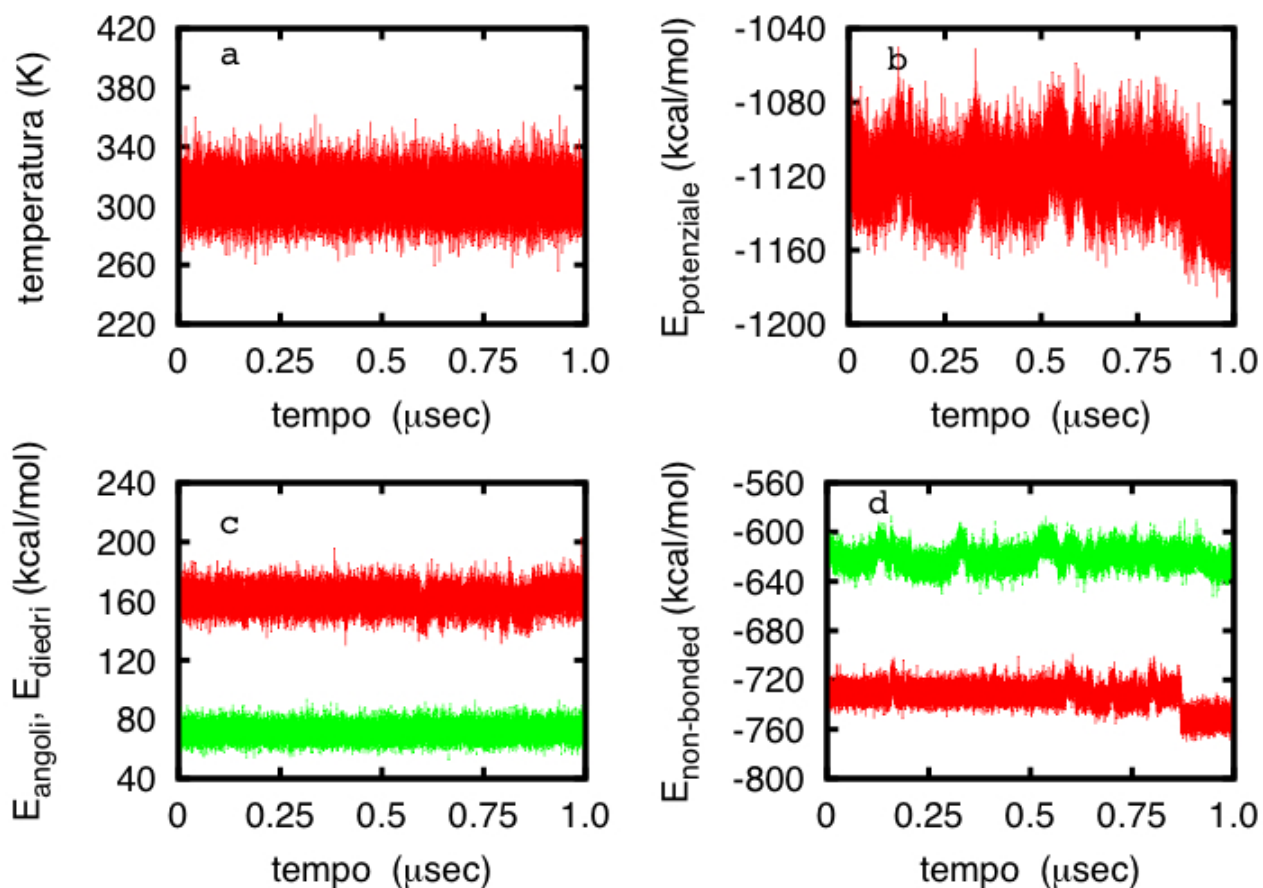


Figura 4.28: Grafici in funzione del tempo della temperatura (a), dell'energia potenziale (b), del potenziale degli angoli (linea rossa in (c)) e dei diedri (linea verde in (c)), e del potenziale dato dai termini non bonded locali (linea rossa in (d)) e dai non bonded non locali (linea verde in (d)) riguardanti la simulazione con Gag e proteasi inizialmente separate, con il campo di forze n^o3 e la struttura di partenza '2457'.

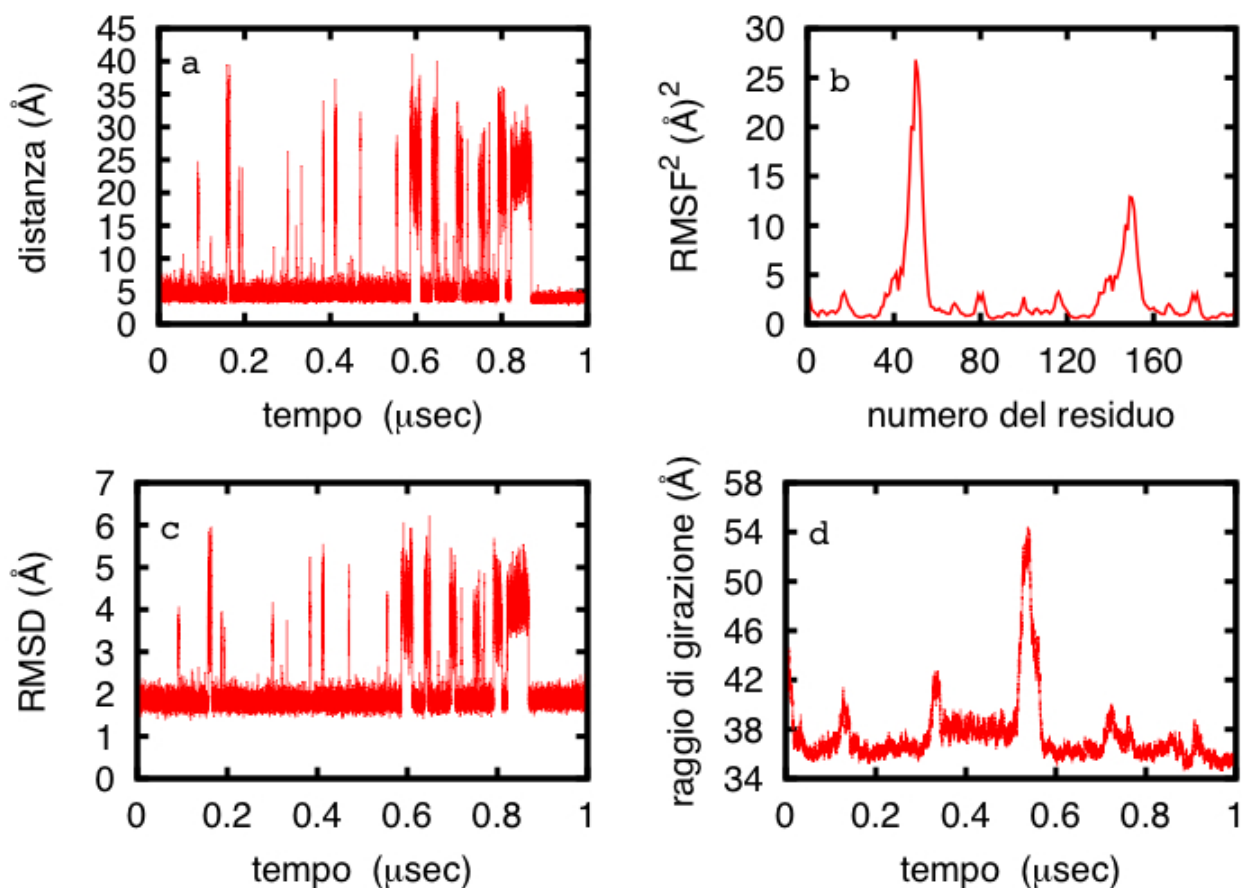


Figura 4.29: Grafici della distanza tra le punte dei flap in funzione del tempo della simulazione della proteasi con la poliproteina Gag (a), di RMSF^2 (b), di RMSD (c), e del raggio di girazione del complesso Gag-Pr al variare del tempo (d) riguardanti la simulazione tra la PR e Gag inizialmente separate, con il campo di forze n°3 e la struttura di partenza '2457'.

ad aggregare con la Gag anche in posizioni diverse rispetto al sito attivo. Questi risultati sono riportati in appendice K.

Interessanti risultano anche le simulazioni a partire dalla proteasi agganciata lungo il linker, ma in posizioni differenti rispetto al sito di taglio, effettuate con il campo n°3. Nelle figure 4.31, 4.32 e 4.33, si vede lo spostamento della proteasi lungo il linker per mezzo della variabile σ rispetto al tempo. In figura 4.31, dove la proteasi è spostata inizialmente di 2 amminoacidi rispetto al sito di taglio, vediamo un immediato corretto riposizionamento della proteasi, mentre in figura 4.32, possiamo vedere un rapido scivolamento di pr lungo il linker per riagganciarsi al sito di taglio. Diversamente accade in figura 4.33, dove la struttura di partenza di Gag è stata ricavata dopo un riassetamento dell'intero sistema Gag-pr. In questo caso la proteasi comincia immediatamente a staccarsi dal linker e successivamente, dopo più di $0.4\mu s$, si aggancia al sito di taglio. I valori delle energie e delle temperature lungo i run, gli RMSF^2 e gli RMSD delle simulazioni con la proteasi spostata lungo il linker, sono anch'essi riportati in appendice K.

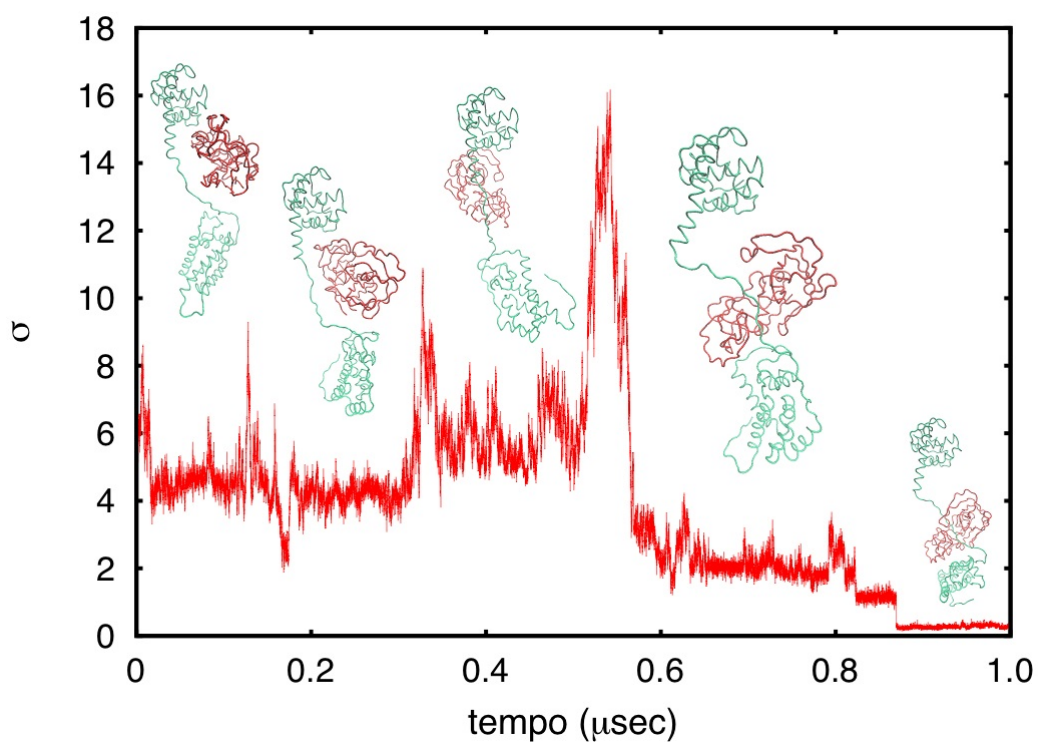


Figura 4.30: Grafico della nuova variabile σ in funzione del tempo mostrandone l'interazione PR-Gag, riguardante la simulazione con Gag e proteasi inizialmente separate, campo di forze n°3 e struttura di partenza '2457'.

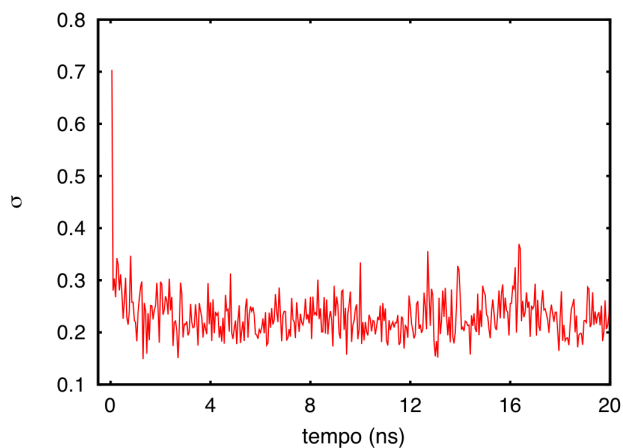


Figura 4.31: Grafico della nuova variabile σ in funzione del tempo nella simulazione in cui pr è spostata inizialmente di 2 amminoacidi rispetto al sito di taglio.

4.5 Conclusioni

È stato così raggiunto lo scopo di questa tesi di analizzare l'interazione che occorre tra l'enzima proteasi e la poliproteina Gag del virus HIV-1. Dai risultati ottenuti dalle varie simulazioni, salta all'occhio come la proteasi impieghi meno tempo ad attaccarsi al sito

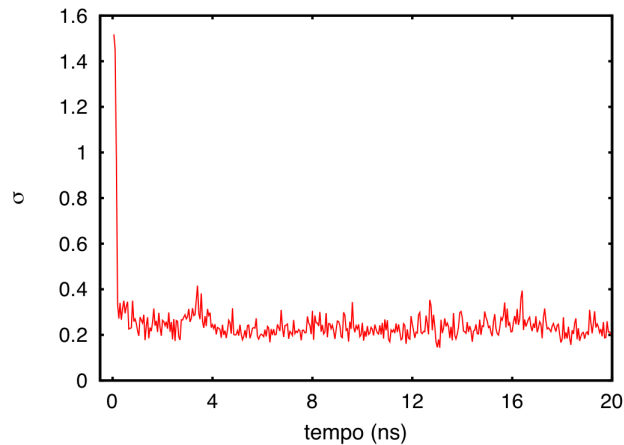


Figura 4.32: Grafico della nuova variabile σ in funzione del tempo nella simulazione in cui pr è spostata inizialmente di 4 amminoacidi rispetto al sito di taglio.

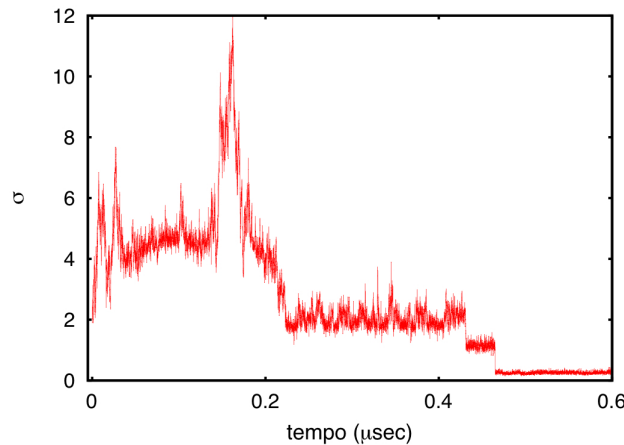


Figura 4.33: Grafico della nuova variabile σ in funzione del tempo nella simulazione in cui pr è spostata inizialmente di 4 amminoacidi rispetto al sito di taglio, e con la struttura di partenza di Gag che è stata ricavata dopo un riassetamento dell'intero sistema Gag-pr.

attivo del complesso MA-CA_{NTD} rispetto a quanto impieghi per la cattura del substrato oligopeptidico. Questo fatto è verosimilmente dovuto all'interazione tra due fattori: uno di natura entropica e l'altro di natura strutturale-dinamica. Da una parte l'inclusione dell'intera Gag nella simulazione e il fissaggio delle sue estremità riduce la mobilità globale del sito di aggancio rispetto al caso in cui nella simulazione è incluso solo l'oligopeptide che lo rappresenta. Questo entropicamente favorisce l'incontro. Tuttavia il tempo di aggancio si riduce anche rispetto alla simulazione in cui l'oligopeptide è tenuto ad estremità fissate. Questo indica che la flessibilità dinamica della parte destrutturata di Gag intorno al sito di aggancio favorisce il processo.

Il corretto posizionamento della sito di taglio nel sito attivo è fondamentale affinché avvenga il processo di taglio della poliproteina da parte dell'enzima, ottenendo così le strutture separate di MA e CA_{NTD} (figura 4.34), che si ricombineranno nella formazione del virione

maturato. Le simulazioni con configurazioni di partenza agganciate ma spostate mostrano che il modello è in grado di riprodurre correttamente anche la selettività per il sito di taglio. Inoltre vi è un'ulteriore conferma dell'importante ruolo svolto dai flap nella dinamica dell'aggancio e della loro asimmetria durante il processo.

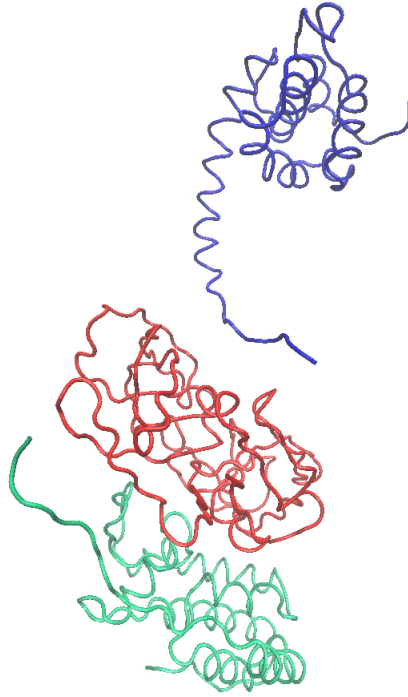


Figura 4.34: Taglio di MA-CA_{NTD} ad opera della proteasi di HIV-1. In blu è rappresentata la struttura MA, in verde CA_{NTD} ed in rosso la proteasi.

Capitolo 5

Conclusioni e possibili sviluppi

In questo lavoro è stato sviluppato e ottimizzato un modello a bassa risoluzione (minimalista) capace di riprodurre nelle simulazioni la struttura e la dinamica delle proteine di HIV-1 pr, Gag e la loro interazione. Il modello prende le fila da uno precedentemente utilizzato per la simulazione dell'interazione tra proteasi e substrati modello (oligopeptidi). In questo lavoro il modello è stato esteso includendo la poliproteina Gag, substrato naturale della proteasi. Questo ha comportato l'ottimizzazione del campo di forze di interazione tra Gag e proteasi, oltre alla creazione de-novo del modello per la Gag, basato su un set di strutture sperimentali ottenute tramite NMR. Questo modello ha un valore di per sé in quanto include la corretta flessibilità della proteina virale, e specialmente della parte destrutturata che contiene il sito di taglio riconosciuto dalla proteasi. La riproduzione della corretta struttura e dinamica di questa parte risulta infatti essenziale per simulare correttamente il processo di aggancio della proteasi.

Il risultato della procedura è un modello per il complesso che è in grado di riprodurre in simulazioni dal modesto costo computazionale il processo di aggancio della proteasi alla Gag, nonché la selettività di questa rispetto al sito di taglio e la corretta struttura e dinamica del complesso. L'esplorazione della dinamica ha rivelato i modi di legame della proteina. Questo potrà essere usato per la ricerca di specifici loci, possibili candidati come bersagli di terapie inibitorie del legame. Poiché l'aggancio della proteasi alla Gag è il passo preliminare dell'idrolisi, e quindi della produzione delle proteine strutturali del virus e della sua maturazione, questa inibizione è una possibile strategia terapeutica contro l'AIDS. Nella pratica, le strutture intermedie ottenute dalle simulazioni di aggancio prodotte in questa tesi, potrebbero essere la base per la progettazione in silico di farmaci inibitori, soprattutto tenuto conto che, mentre le strutture termodinamicamente stabili sono più facilmente accessibili sia all'esperimento che a tecniche modellistiche standard, strutture di transizione ed intermedi sono generalmente elusivi. In questo lavoro è stato tuttavia analizzato solo il passo base del processo, e solo una porzione spazialmente ristretta del sistema, includente solo un'unità base (una sola Gag e una sola proteasi). È noto tuttavia che nel virione immaturo le Gag sono organizzate in un reticolo bidimensionale, e naturalmente una quantità statisticamente macroscopica di proteasi è presente. Quindi un'applicazione immediata dei risultati qui ottenuti consiste nel simulare una porzione macroscopica realistica del virus, disponendo le Gag nel modo in cui si trovano all'interno del virione e introducendo nella simulazione un numero sufficiente di proteasi. Questo comporta solamente aumentare il numero di molecole pre-

senti nel sistema attuale, lasciando invariato il modello, ed è numericamente realizzabile su sistemi di calcolo (anche non massicciamente) paralleli. Questa sarebbe la prima simulazione dinamica su scala macroscopica dell'interno del virus di HIV.

Con graduali modifiche successive del modello sarebbe possibile poi analizzare i passi successivi, come il riarrangiamento dei frammenti di Gag idrolizzati ed eventualmente la formazione del capsido. Maggiore la dimensione del sistema, maggiori le risorse di calcolo necessarie. Tuttavia, data la semplicità dei modelli, anche l'intero capsido del virus sarebbe accessibile su sistemi ad alte prestazioni. Anche per i passi successivi della maturazione, una conoscenza dinamica dei sistemi in esame potrebbe portare a chiarire alcuni aspetti ancora oscuri del processo di maturazione e dare la possibilità di interagire con essi, precisamente fornendo indicazioni utili per la progettazione di nuovi inibitori, come stati strutturali transitori difficilmente accessibili se non con tecniche di simulazione al calcolatore. Il motivo ultimo che spinge questa ricerca è dare un contributo al controllo e possibilmente eliminazione dell'AIDS, che, in alcune parti del mondo è diffuso al livello epidemico, e rischia di diffondersi globalmente al livello endemico.

Appendice A

Tabella amminoacidi

Amminoacido	codice a tre lettere	codice ad una lettera
Alanina	Ala	A
Arginina	Arg	R
Asparagina	Asn	N
Acido aspartico	Asp	D
Cisteina	Cys	C
Glicina	Gly	G
Acido glutammico	Glu	E
Glutammina	Gln	Q
Istidina	His	H
Isoleucina	Ile	I
Leucina	Leu	L
Lisina	Lys	K
Metionina	Met	M
Fenilalanina	Phe	F
Prolina	Pro	P
Serina	Ser	S
Treonina	Thr	T
Triptofano	Trp	W
Tirosina	Tyr	Y
Valina	Val	V

Tabella A.1: Tabella indicante i nomi degli amminoacidi nella prima colonna e la relativa nomenclatura nelle restanti due colonne in base al codice a tre lettere o ad una lettera.

Appendice B

Risultati dei lavori sulla predizione delle sequenze idrolizzate dalla proteasi

Nello studio di Chou [21] sono stati presi due set di dati, un set positivo contenente i peptidi tagliati dalla proteasi (chiamato S^+) e l'altro contenente i peptidi non tagliati dalla proteasi (S^-). Per ogni amminoacido è stata calcolata la frequenza con cui si presenta nei vari sottositi per entrambi i set, ad esempio $P^+(X_i)$ rappresenta la probabilità che l'amminoacido X_i occorra al sottosito R_i per quanto riguarda il set positivo. Inoltre è stata calcolata, sempre per entrambi i set, la probabilità condizionale $P_j(X_i|X_j)$ che un amminoacido X_i occorra alla posizione R_i dato l'amminoacido X_j nella posizione R_j . Così è stata introdotta la probabilità di un dato peptide di appartenere al set positivo:

$$\begin{aligned} \psi^+(X_4X_3X_2X_1X_{1'}X_{2'}X_{3'}X_{4'}) = & P_4^+(X_4) + P_3^+(X_3|X_4) + P_2^+(X_2|X_3) + \\ & P_1^+(X_1|X_2) + P_{1'}^+(X_{1'}|X_1) + P_{2'}^+(X_{2'}|X_{1'}) + \\ & P_{3'}^+(X_{3'}|X_{2'}) + P_{4'}^+(X_{4'}|X_{3'}) \end{aligned}$$

e quella di appartenere al set negativo:

$$\begin{aligned} \psi^-(X_4X_3X_2X_1X_{1'}X_{2'}X_{3'}X_{4'}) = & P_4^-(X_4) + P_3^-(X_3|X_4) + P_2^-(X_2|X_3) + \\ & P_1^-(X_1|X_2) + P_{1'}^-(X_{1'}|X_1) + P_{2'}^-(X_{2'}|X_{1'}) + \\ & P_{3'}^-(X_{3'}|X_{2'}) + P_{4'}^-(X_{4'}|X_{3'}) \end{aligned}$$

Se per un peptide si ha che $\psi^+ > \psi^-$ allora può essere tagliato dalla proteasi, altrimenti no. Questo metodo si chiama algoritmo della funzione discriminante ed è basato sulla teoria delle catene di Markov, in modo che sono inclusi sia gli effetti tra i sottositi con posizioni adiacenti che quelli con posizioni alternate. Quindi i siti di taglio della proteasi di HIV in una proteina sono prevedibili dalla sua struttura primaria.

In [35] viene mostrato che il problema della specificità del sito di taglio presente nei substrati da parte della proteasi di HIV è un problema linearmente separabile e l'applicare

a questo problema dei classificatori non lineari è inutile, in quanto grazie ai classificatori lineari si possono avere buone predizioni come dagli algoritmi non lineari.

Nella classificazione lineare si determina un iperpiano $\{x|\langle w, x \rangle + b = 0\}$ in $X = \mathfrak{R}^n$ in modo che un insieme di punti dati P stia nel semispazio positivo delimitato da esso e un altro insieme di punti N stia nel semispazio negativo: in questo caso allora $f = \text{sgn}\{\langle w, x \rangle + b\}$ [66].

Un classificatore non lineare invece trova una relazione tra f e x della forma $f = \phi[w, x]$ in $X = \mathfrak{R}^n$.

E' stato mostrato che il campione completo della proteasi di HIV-1 è linearmente separabile: esiste almeno un iperpiano che separa gli ottameri tagliati da quelli non tagliati.

Il secondo esperimento del lavoro descritto in [35] è stato quello di rimuovere delle posizioni P_i dagli ottameri P e vedere se il problema rimaneva lineare. Si è trovato che le posizioni $P3$ e $P3'$ o $P3$ e $P4'$ potevano essere rimosse senza cambiare la separabilità lineare dei dati. Inoltre le posizioni più importanti per la scissione erano $P2$, $P1$, $P1'$ e $P2'$, e sono state trovate le seguenti regole che aumentano la probabilità di taglio dell'octapeptide:

- un acido glutammico (E) o una glutammina (Q) nella posizione $P2'$;
- una fenilalanina (F), una tirosina (Y) o una leucina (L) nella posizione $P1$;
- una prolina (P) in posizione $P1'$;
- una valina (V) o un'alanina (A) in posizione $P2$.

Le regole elencate hanno senso da un punto di vista chimico. Un acido glutammico (E) o una glutammina (Q) sono amminoacidi simili, scambiare uno con l'altro nella posizione $P2'$ non deve fare una grande differenza. La fenilalanina (F), la tirosina (Y), la metionina (M) e la leucina (L) sono anche loro amminoacidi simili e ha senso scambiarli. Allo stesso modo, la valina (V) e l'alanina (A) sono molto simili e lo scambio nella posizione $P2$ non dovrebbe cambiare molto.

Un altro metodo per la predizione dei siti di taglio, proviene dai due problemi di scissione, la scissione di corti oligopeptidi e la scissione delle proteine native, i quali sono correlati ma differenti. I brevi oligopeptidi e le proteine denaturate non hanno strutture foldate. Le scissioni delle proteine native sono casi rari, il che implica che i siti di scissione sono in una piccola regione della sequenza proteica. Liwen You, nell'articolo [36], cerca di predire i potenziali siti di taglio nelle proteine native non virali combinando le informazioni ottenute da brevi oligopeptidi, strutture secondarie e dall'accessibilità del solvente. Un predittore basato solo su brevi oligopeptidi, non funziona bene sulle proteine native dato che alcuni previsti siti di taglio potrebbero non essere esposti alla proteasi o le strutture secondarie locali potrebbero impedire il legame con la proteasi.

In [37] è spiegata la costruzione di un modello per la predizione dei siti di taglio della proteasi di HIV-1 basato sulle proprietà fisico-chimiche e biologiche degli amminoacidi. Queste proprietà sono state divise in tre gruppi: struttura secondaria, idrofobicità ed altre. E' stato trovato che le posizioni $P1$ and $P2'$ degli amminoacidi dell'octapeptide contribuiscono maggiormente alla specificità del substrato della proteasi di HIV-1 rispetto agli altri siti. In $P1$ si preferiscono i residui idrofobici, mentre gli amminoacidi che occorrono in $P2'$ sono

maggiormente determinati dalle proprietà delle strutture secondarie. Inoltre i risultati sono consistenti con il fatto che i residui idrofobici sono preferiti in P_1 e P'_1 .

Nel lavoro [38] è presentato un metodo che si basa sull'ipotesi che certe posizioni dovrebbero contenere determinati amminoacidi. Così vengono stabilite delle regole per trovare le specificità delle scissioni dei substrati. Le regole trovate per le posizioni P_1 e P'_1 sono le regole più consistenti e in accordo con i precedenti risultati per la proteasi di HIV-1: nella posizione P_1 sono preferiti grandi residui idrofobici e nella posizione P'_1 sono preferiti residui idrofobici.

Ogul ha studiato un altro modello [39] dove la sua importanza sta nel fatto che può includere le interazioni tra i residui della sequenza del sito di taglio poichè l'ordine degli amminoacidi in un peptide ha un'influenza nel legame con la proteasi.

In un altro articolo [40] è stato utilizzato un metodo con una classificazione non lineare per cercare il minimo numero di caratteristiche rilevanti che descrivevano il set di dati.

Nanni e Lumini [41] hanno proposto un nuovo descrittore di peptidi o proteine ottenuto da una rappresentazione matriciale, che è trattata come un'immagine. Una matrice deve contenere le informazioni sulla posizione degli amminoacidi nella sequenza e anche le loro proprietà fisico-chimiche. Le relazioni tra tutte le coppie di amminoacidi che costituiscono la sequenza del peptide o della proteina sono raccolte in un'altra matrice e la matrice finale ottenuta è riscalata in livelli di grigio e visualizzata come un'immagine.

Appendice C

DL_POLY

All'interno di DL_POLY l'energia potenziale di configurazione di un sistema molecolare può assumere la forma generale

$$\begin{aligned} U(r_1, r_2, \dots, r_N) = & \sum_{i_{bond}=1}^{N_{bond}} u_{bond}(i_{bond}, r_a, r_b) \\ & + \sum_{i_{angle}=1}^{N_{angle}} u_{angle}(i_{angle}, r_a, r_b, r_c) \\ & + \sum_{i_{dihedral}=1}^{N_{dihedral}} u_{dihedral}(i_{dihedral}, r_a, r_b, r_c, r_d) \\ & + \sum_{i_{inv}=1}^{N_{inv}} u_{inv}(i_{inv}, r_a, r_b, r_c, r_d) \\ & + \sum_{i=1}^{N-1} \sum_{j>i}^N u_{pair}(i, j, |r_i - r_j|) \\ & + \sum_{i=1}^{N-2} \sum_{j>i}^{N-1} \sum_{k>j}^N u_{3body}(i, j, k, r_i, r_j, r_k) \\ & + \sum_{i=1}^{N-1} \sum_{j>i}^N u_{Tersoff}(i, j, r_i, r_j, R^N) \\ & + \sum_{i=1}^{N-3} \sum_{j>i}^{N-2} \sum_{k>j}^{N-1} \sum_{n>k}^N u_{4body}(i, j, k, n, r_i, r_j, r_k, r_n) + \end{aligned}$$

$$\begin{aligned}
& + \sum_{i=1}^N u_{Metal}(i, r_i, R^N) \\
& + \sum_{i=1}^N u_{extn}(i, r_i, v_i)
\end{aligned} \tag{C.1}$$

dove u_{bond} , u_{angle} , u_{dihed} , u_{inv} , u_{pair} , u_{3body} , $u_{Tersoff}$ e u_{4body} sono funzioni empiriche di interazione che rappresentano rispettivamente legami chimici, angoli di valenza, angoli diedri, angoli di inversione, interazioni a due e a tre corpi, Tersoff (legami covalenti di molti corpi) e forze di interazioni a quattro corpi. I primi quattro termini sono considerati come interazioni intramolecolari mentre i successivi cinque come intermolecolari. Il termine u_{metal} è un potenziale metallico dipendente dalla densità ed il termine finale u_{extn} rappresenta un potenziale di campo esterno. I vettori r_a , r_b , r_c e r_d si riferiscono alle posizioni degli atomi coinvolti in una data interazione. Un vettore speciale R^N è usato per indicare una dipendenza tra molti corpi. I numeri N_{bond} , N_{angle} , N_{dihed} e N_{inv} si riferiscono ai numeri totali delle rispettive interazioni presenti nel sistema simulato, e gli indici i_{bond} , i_{angle} , i_{inv} e i_{dihed} specificano univocamente un'interazione individuale di ciascun tipo. Naturalmente, a seconda del campo di forze che si vuole implementare, uno o più dei termini possono essere nulli. In particolare nel campo di forze del modello qui usato i termini presenti sono u_{bond} , u_{angle} , $u_{dihedral}$ ed u_{pair} .

Paragonando i termini dell'eq. 3.22 con quelli dell'eq. C.1 si evince che U_b è dato dalla somma dei vari u_{bond} , U_θ dalla somma degli u_{angle} ed U_ϕ da quella degli $u_{dihedral}$. Quindi $u_{bond}(i_{bond}, r_a, r_b)$ definisce il legame chimico i -esimo tra due corpi legati tra loro, $u_{angle}(i_{angle}, r_a, r_b, r_c)$ il potenziale dato dall'angolo i -esimo formato da tre corpi legati tra loro e $u_{dihedral}(i_{dihedral}, r_a, r_b, r_c, r_d)$ il potenziale dato dall' i -esimo diedro formato da quattro corpi legati tra loro. u_{pair} rappresenta le forze di van der Waals, i potenziali elettrostatici ed eventuali altre forze non descrivibili come legami chimici (ad esempio interazioni idrofobiche, π stacking, legami a idrogeno) che nell'eq.3.22 sono inclusi nel termine U_{nb} .

Per eseguire DL_POLY sono necessari tre file principali: CONTROL, FIELD e CONFIG. Questi file devono essere presenti nella directory da cui viene eseguito il programma. L'esecuzione del programma produce i seguenti file nella directory di esecuzione: OUTPUT, REVCON e REVIVE (e facoltativamente STATIS, HISTORY, RFDAT and ZDNDAT). Il file CONTROL contiene i parametri di controllo della simulazione: numero e ampiezza dei timestep, temperatura (e tipo di termostati), pressione, tipo di dati scritti nell'output e frequenza di scrittura, etc. Il file CONFIG contiene le posizioni degli atomi e opzionalmente anche le velocità. Il file FIELD specifica la natura delle interazioni intermolecolari, la topologia molecolare e le proprietà atomiche, come la carica e la massa. Tra il file FIELD e il file CONFIG c'è una corrispondenza strutturale. È necessario che l'ordine di specificazione dei tipi molecolari e dei loro componenti atomici nel file FIELD segua l'ordine in cui appaiono nel file di CONFIG.

Tra i file di dati generati dall'esecuzione di DL_POLY, il file OUTPUT fornisce una sintesi effettiva del lavoro di esecuzione: le informazioni in ingresso, la configurazione di partenza, i dati termodinamici istantanei e mediati, le configurazioni finali, le funzioni di

distribuzione radiale (RDFs) e i dati del tempo di lavoro. Il file di riavvio REVIVE contiene i dati accumulati per un numero di quantità termodinamiche e RDFs, ed è destinato ad essere utilizzato come file di input per un ciclo successivo. Il file REVCAN contiene la configurazione di riavvio, vale a dire le ultime posizioni, velocità e forze degli atomi quando l'esecuzione finisce. Il file STATIS contiene un elenco di valori istantanei di termodinamica e di altre variabili, in una forma adatta per l'analisi temporale o statistica. Infine il file HISTORY fornisce una sequenza temporalmente ordinata di configurazioni per agevolare il proseguimento dell'analisi dei moti atomici.

In figura C.1 è rappresentato uno schema contenente i file di entrata e quelli prodotti dal pacchetto DL.POLY.

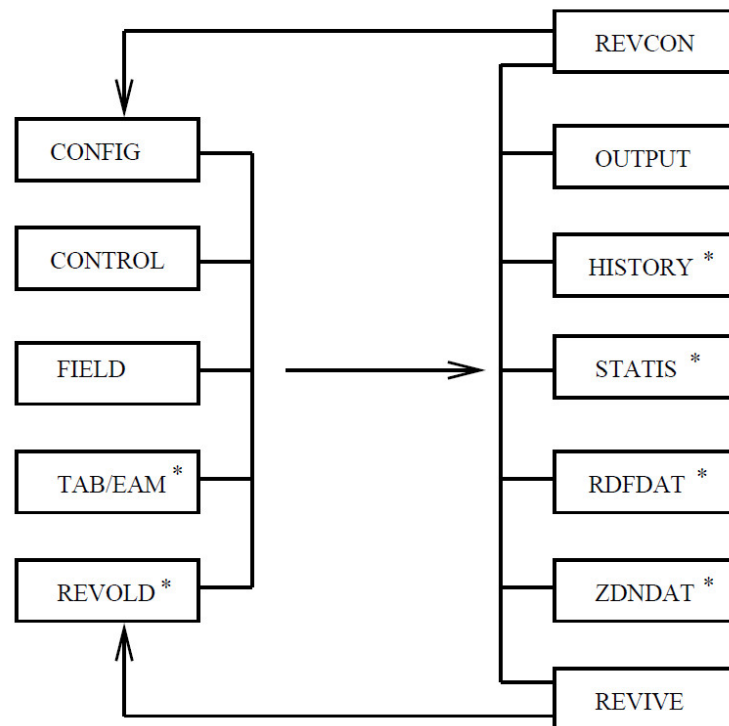


Figura C.1: Schema rappresentante i file di DL.POLY di ingresso, posti a sinistra, e i file generati in uscita, posti a destra. Gli asterischi si riferiscono ai file non obbligatoriamente presenti nella generazione di una simulazione [47].

Appendice D

Parametri del campo di forze

valore di equilibrio	k	commenti
legami		
3.8 Å	SHAKE ¹	usato nei vicoli che mantengono fisse le distanze
angoli		
$\theta_\alpha = 90$ gradi	38 kcal/mole rad ²	valore di k ottimizzato
θ_β : dipende dagli amminoacidi, vedi Tabella D.2	17 kcal/mole rad ²	valore di k ottimizzato
diedri		
ref. struttura (1HHP)	5 kcal/mole rad ²	

Tabella D.1: Parametri delle interazioni bonded.

¹L'algoritmo SHAKE [47] usato per i legami che mantengono le distanze ad un valore fissato, si basa sull'integratore Verlet Leapfrog ed è un processo a due passi. Nel primo si calcola il moto dei siti del sistema assumendo che non vi siano legami rigidi. Nel secondo passo viene reintrodotta la distanza fissa tra i bead: dalla variazione del legame rigido si calcola la forza necessaria per la conservazione della lunghezza del legame.

Amminoacido			θ_β (gradi)
X	Ala	X	130
Gly	Ala	Asp,Thr	120
Ile,Leu	Ala	Lys,Glu	135
X	Asn	X	120
Met,Ile	Asn	Leu,Val	110
Leu,Ile	Asn	Glu,Lys	135
X	Gln	X	120
Tyr,Pro,Pro	Gln	Arg,Ile,Val	115
Gly	Gln	Leu	135
X	Glu	X	120
Lys,Leu,Ile,Met	Glu	Ala,Glu,Ile,Glu	115
X	Gly	X	135
Ile,Pro	Gly	Thr,Val	150
X	His	X	130
Gly	His	Lys	120
X	Ile	X	120
Cys,Ile,Gly	Ile	Gly,Glu	110
Gly,Gly,Ile,Pro,Thr	Ile	Leu,Lys,Met,Asn,Lys	135
X	Leu	X	120
Asn,Leu,Leu,Leu,Trp	Leu	Thr,Asp,Asn,Thr,Thr	100
Glu,Pro,Pro	Leu	Val,Asn,Ser	135
X	Met	X	120
Asn,Ser	Met	Glu,Glu	140
X	Phe	X	120
Ile	Phe	Gly	135
X	Ser	X	120
Leu	Ser	Met	115
X	Thr	X	120
Val,Val,Val,Val,Ile,Pro	Thr	Asp,Gly,Ile,Leu,Leu,Pro	115
Leu	Thr	Ala,Cys	135
X	Trp	X	120
Lys	Trp	Lys,Arg	135
X	Val	X	120
Arg,Pro,Leu	Val	Lys,Asn,Thr	135
X	Asp	X	130
X	Cys,Lys,Tyr	X	120
X	Arg,Pro	X	115

Tabella D.2: Valori di equilibrio di θ_β per tutte le triplette di amminoacidi (X=qualsiasi amminoacido).

Appendice E

Interazioni non bonded locali Pr-substrato

aa Pr	aa sito di taglio	r_0 (Å)
25	5	7.920
25	6	7.028
27	3	7.452
27	4	6.134
27	5	5.341
27	6	7.102
28	3	5.770
28	4	3.696
28	5	5.456
28	6	7.009
46	1	7.786
47	1	6.629
47	2	5.641
47	3	6.169
47	4	7.076
48	1	7.308
48	2	6.443
48	3	5.084
48	4	6.284
48	5	7.923
49	3	5.527
49	4	5.147
49	5	4.984
49	6	7.243
50	5	6.029
50	6	7.553
50	7	7.158

Tabella E.1: continua nella prossima pagina

Tabella E.1: continua dalla pagina precedente

aa Pr	aa sito di taglio	r_n^0 (Å)
126	5	6.810
126	6	4.719
126	7	6.355
127	5	7.084
127	6	5.217
127	7	4.261
127	8	6.053
128	7	6.088
128	8	5.301
128	9	7.309
129	7	7.070
129	8	6.723
129	9	7.349
144	9	7.925
145	9	6.363
146	7	7.027
146	8	5.817
146	9	4.233
147	7	5.750
147	8	5.656
147	9	6.217
148	5	6.762
148	6	5.382
148	7	4.649
148	8	6.600
149	4	7.302
149	5	7.280
149	6	6.761
149	7	7.846

Tabella E.1: Valori della distanza di equilibrio r_0 (terza colonna) delle interazioni non bonded locali tra gli amminoacidi della proteasi (prima colonna) e quelli del sito di taglio (seconda colonna).

Appendice F

Tabelle valori RMSD

n° struttura	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	0	2.036	2.544	1.471	2.354	2.177	2.695	2.185	2.930	1.970	2.300	2.398	1.816	1.639	2.488	2.903	1.664	2.236	2.119	1.331
1	0	1.611	2.698	2.751	2.645	2.077	2.481	2.006	2.386	2.512	2.768	2.255	1.892	1.631	2.030	1.750	2.750	2.990	1.966	1.966
2			0	2.830	2.382	2.456	1.489	2.378	1.695	2.247	2.002	2.362	2.024	2.631	1.714	1.473	2.156	2.579	3.079	2.428
3				0	1.867	2.389	2.972	2.318	3.303	2.204	2.266	2.433	1.939	2.019	2.713	2.946	2.085	2.341	2.224	1.317
4					0	1.511	2.120	1.362	2.490	1.400	1.265	1.378	1.832	2.365	2.537	2.199	2.299	1.479	1.766	2.218
5						0	2.099	1.508	2.453	1.200	1.118	1.284	1.915	2.448	2.818	2.574	2.374	0.797	1.310	2.570
6							0	1.789	1.272	1.592	1.632	1.689	1.661	2.366	1.504	1.287	2.060	2.275	2.776	2.625
7								0	1.952	0.932	1.389	1.252	1.657	2.075	2.426	2.086	2.010	1.405	1.605	2.373
8									0	2.125	2.279	2.308	2.144	2.447	1.830	1.249	1.933	2.445	2.874	2.812
9										0	1.009	0.844	1.408	1.997	2.285	2.204	2.066	1.321	1.711	2.243
10											0	1.014	1.680	2.355	2.385	2.098	2.392	1.332	1.864	2.375
11												0	1.540	2.486	2.442	2.257	2.451	1.412	1.938	2.562
12													0	2.064	1.904	1.979	1.771	1.955	2.205	1.792
13														0	1.895	2.374	1.334	2.468	2.385	1.399
14															0	1.401	1.651	2.959	3.284	2.004
15																0	1.911	2.588	2.959	2.461
16																	0	2.347	2.346	1.629
17																		0	1.024	2.568
18																			0	2.549
19																				0

Figura F.1: RMSD calcolato da *Gly1* a *Thr121* per le 20 strutture aventi codice pdb 1L6N

n° struttura	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	0	8.704	9.231	10.419	7.580	10.487	7.270	12.213	7.981	6.864	9.394	8.233	8.335	9.068	8.936	8.303	7.314	7.768	12.120	12.074
1		0	16.512	17.561	13.954	17.957	7.506	19.630	14.502	8.513	14.369	14.313	15.149	13.996	15.957	14.665	14.217	14.583	19.010	18.984
2			0	3.590	6.365	3.535	13.930	5.289	6.034	11.752	8.556	7.129	5.702	6.658	4.909	5.934	5.548	4.108	4.480	4.378
3				0	7.072	2.920	15.428	4.564	5.350	12.619	8.052	7.459	4.983	7.742	5.838	4.913	5.835	6.143	3.594	4.034
4					0	6.584	13.039	7.265	3.726	7.444	8.102	1.740	3.762	6.770	2.965	5.854	3.127	4.693	7.192	6.885
5						0	15.605	3.826	6.031	12.638	8.701	7.010	5.409	8.266	4.901	6.095	5.821	5.611	3.282	3.696
6							0	17.159	13.540	10.071	11.910	13.628	14.447	11.314	14.059	13.179	13.568	12.059	17.010	16.987
7								0	6.501	13.303	9.399	7.166	6.111	8.548	5.119	7.155	6.832	6.484	3.247	3.021
8									0	8.589	7.043	3.681	2.622	8.081	4.152	4.004	3.408	5.798	6.290	6.194
9										0	10.687	7.257	9.320	9.146	9.611	10.088	8.328	9.398	13.083	12.987
10											0	8.263	8.144	7.035	7.940	5.765	8.578	8.268	9.060	9.033
11												0	3.953	6.559	3.385	6.193	3.601	5.269	7.356	7.055
12													0	9.182	3.971	4.232	2.566	5.856	5.951	5.795
13														0	6.369	7.930	8.431	5.116	8.208	8.736
14															0	5.816	4.107	3.683	5.377	5.140
15																0	5.316	6.128	6.737	6.594
16																	0	5.472	6.666	6.297
17																		0	6.495	6.501
18																			0	2.289
19																				0

Figura F.2: RMSD calcolato da *Gly122* a *Pro148* per le 20 strutture aventi codice pdb IL6N

n° struttura	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	0	1.765	1.266	1.325	1.954	1.889	1.284	1.512	1.464	1.540	2.340	1.442	1.817	1.896	1.587	1.388	1.292	1.727	2.570	1.985
1		0	2.152	2.353	2.091	1.061	1.904	1.006	1.199	1.169	1.394	2.159	2.367	1.183	1.336	1.350	1.880	1.163	1.713	1.162
2			0	1.188	2.671	2.229	1.810	2.113	1.963	2.169	2.854	1.932	2.398	2.363	2.193	1.712	1.309	2.143	2.875	2.541
3				0	2.614	2.520	1.606	2.192	2.080	2.242	3.007	1.814	2.348	2.412	2.217	1.783	1.295	2.421	3.028	2.563
4					0	2.268	1.566	1.771	1.857	1.761	1.837	1.619	1.296	2.050	2.111	1.986	2.247	2.156	2.606	2.087
5						0	2.163	1.124	1.208	1.408	1.478	2.263	2.528	1.568	1.428	1.375	1.934	0.706	1.453	1.271
6							0	1.623	1.730	1.722	2.365	1.287	1.454	2.072	1.773	1.548	1.486	2.003	2.669	2.003
7								0	1.014	1.208	1.514	1.889	2.094	1.401	1.284	1.295	1.715	0.989	1.627	1.209
8									0	1.376	1.620	1.836	1.972	1.315	1.358	1.401	1.743	1.101	1.601	1.421
9										0	1.400	1.916	2.142	1.395	1.320	1.433	1.770	1.384	2.096	1.135
10											0	2.415	2.367	1.453	1.869	1.879	2.471	1.585	1.852	1.415
11												0	1.473	1.984	1.993	1.648	1.727	2.023	2.761	2.305
12													0	2.160	2.145	2.100	2.210	2.329	2.966	2.505
13														0	1.539	1.560	2.081	1.564	2.047	1.396
14															0	1.321	1.642	1.296	2.099	1.256
15																0	1.194	1.255	1.989	1.497
16																	0	1.822	2.502	2.031
17																		0	1.488	1.356
18																			0	1.862
19																				0

Figura F.3: RMSD calcolato da Arg149 a Ser280 per le 20 strutture aventi codice pdb IL6N

Appendice G

Valori degli angoli e dei diedri delle strutture NMR

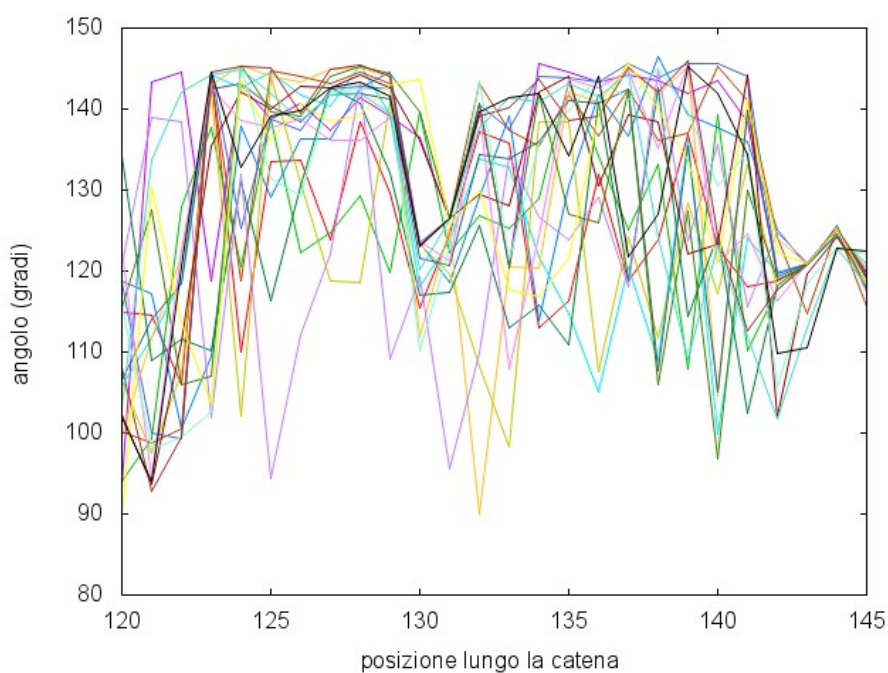


Figura G.1: Valori angoli per le 20 strutture aventi codice pdb 1L6N. La posizione i lungo la catena rappresenta il primo amminoacido dei tre necessari a formare un angolo $(i, i + 1, i + 2)$.

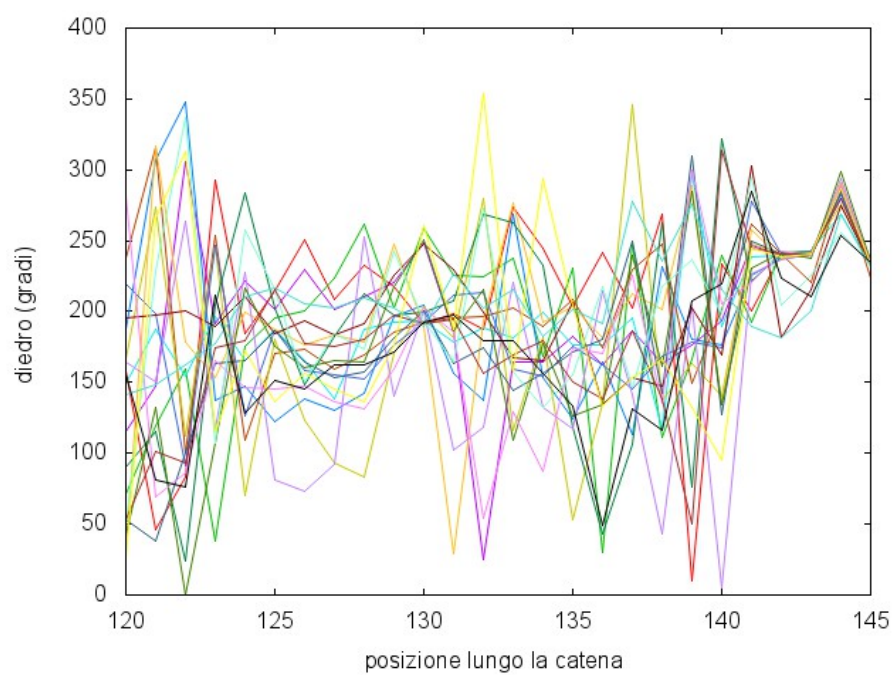
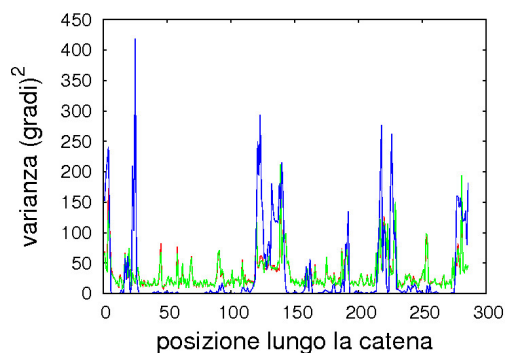


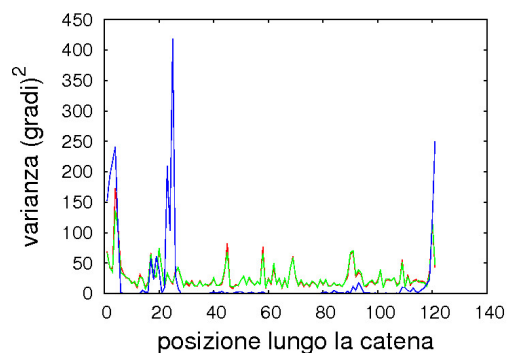
Figura G.2: Valori diedri per le 20 strutture aventi codice pdb 1L6N. La posizione i lungo la catena rappresenta il primo amminoacido dei quattro necessari a formare un angolo diedro $(i, i + 1, i + 2, i + 3)$.

Appendice H

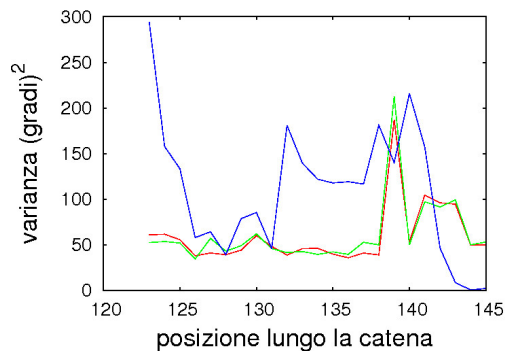
Valori degli angoli di $MA-CA_{NTD}$



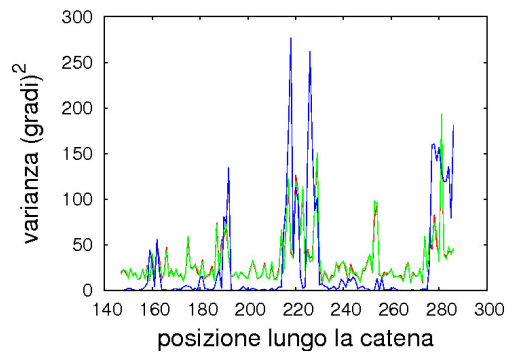
(a) Valori della varianza degli angoli θ in funzione della posizione lungo la catena per $MA-CA_{NTD}$.



(b) Valori della varianza degli angoli θ in funzione della posizione lungo la catena per MA .



(c) Valori della varianza degli angoli θ in funzione della posizione lungo la catena per il linker.



(d) Valori della varianza degli angoli θ in funzione della posizione lungo la catena per CA .

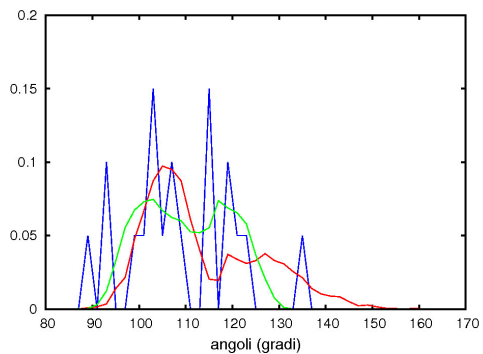
Figura H.1: Confronto tra la simulazione avente la struttura 2 come configurazione di partenza (rosso), quella avente la struttura 13 (verde) e le strutture NMR 1L6N (blu).

Appendice I

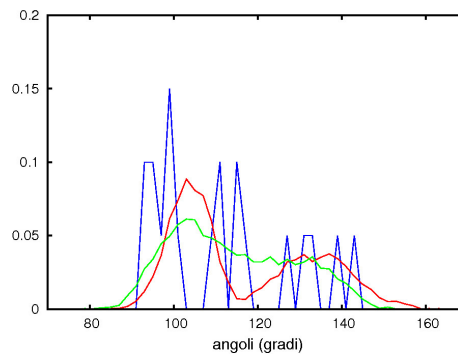
Distribuzioni degli angoli del linker

Il modello bistabile è stato applicato agli angoli del linker aventi le seguenti posizioni lungo la catena: 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142.

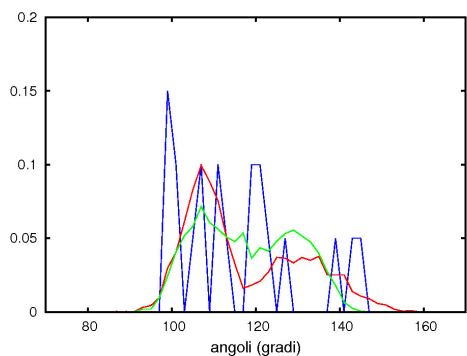
Il potenziale quartico è stato applicato agli angoli del linker aventi le seguenti posizioni lungo la catena: 120, 121, 122, 123, 124, 125, 127, 130, 133, 134, 135, 136, 137, 138, 139, 141, 142.



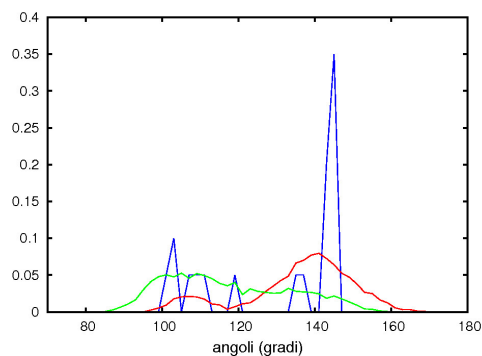
(a) *Posizione 120 lungo la catena.*



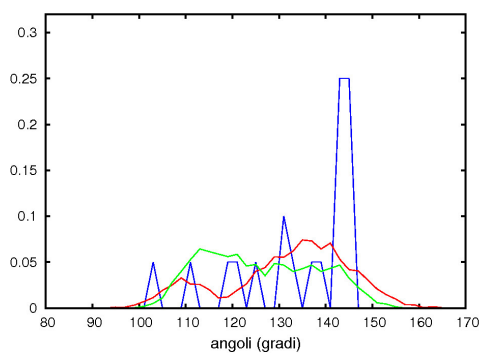
(b) *Posizione 121 lungo la catena.*



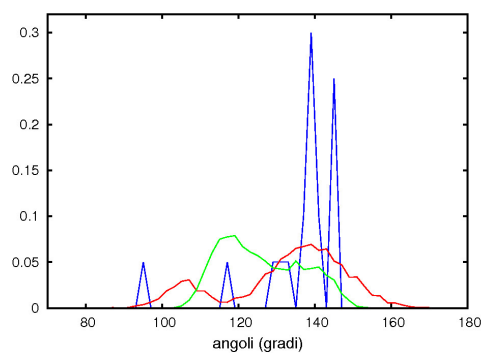
(c) *Posizione 122 lungo la catena.*



(d) *Posizione 123 lungo la catena.*

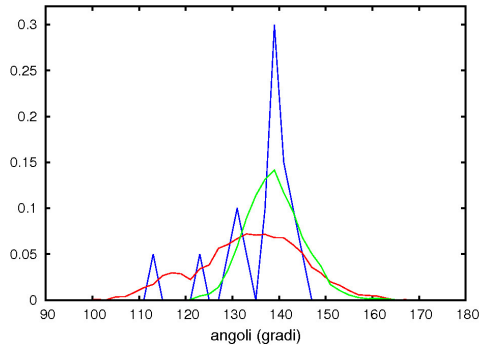


(e) *Posizione 124 lungo la catena.*

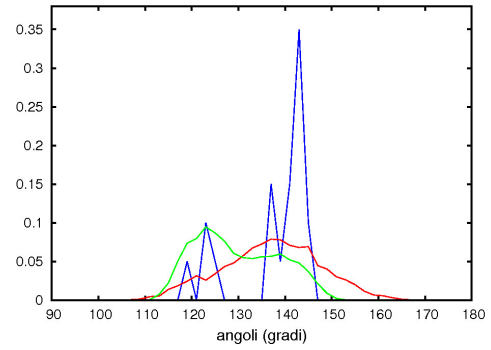


(f) *Posizione 125 lungo la catena.*

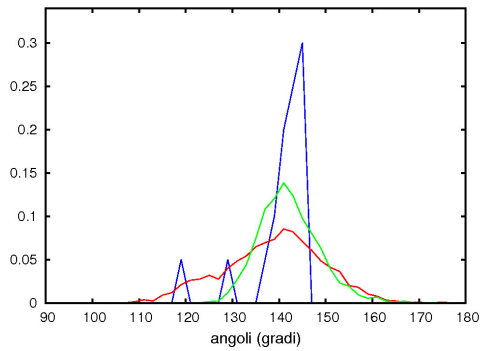
Figura I.1: Distribuzioni normalizzate degli angoli del linker. La linea rossa rappresenta i valori ottenuti dalla simulazione usando il modello bistabile implementato sia sugli angoli che sui diedri del linker, la linea verde i dati ricavati dalla simulazione con il potenziale quartico applicato agli angoli del linker, e la linea blu i dati delle strutture NMR.



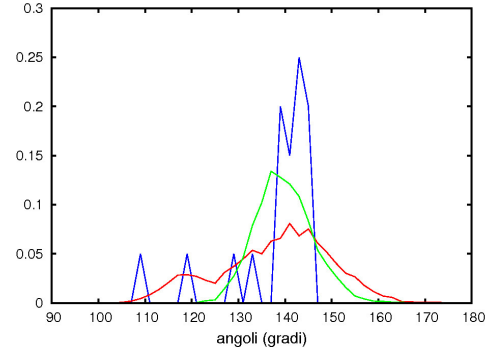
(a) Posizione 126 lungo la catena.



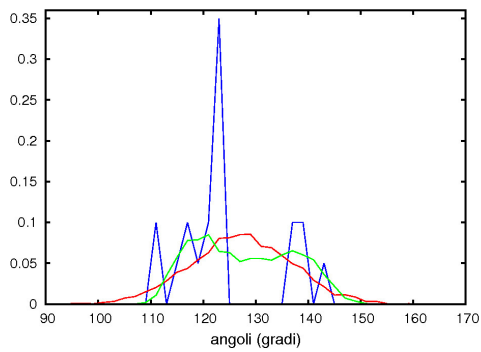
(b) Posizione 127 lungo la catena.



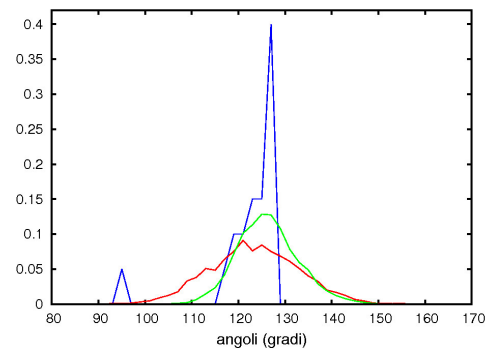
(c) Posizione 128 lungo la catena.



(d) Posizione 129 lungo la catena.

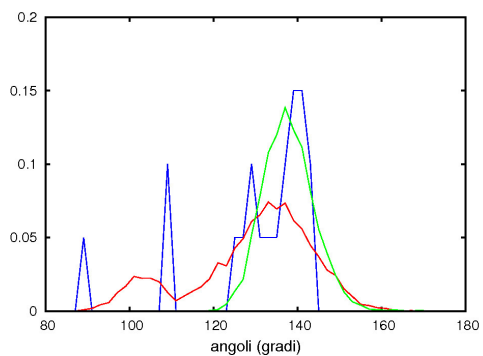


(e) Posizione 130 lungo la catena.

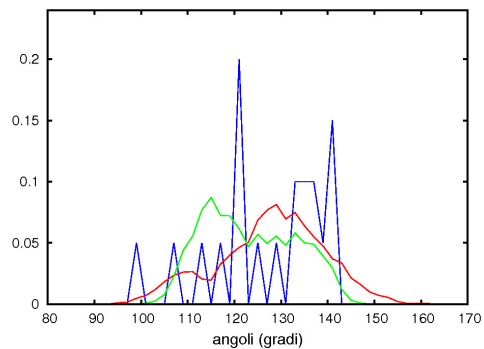


(f) Posizione 131 lungo la catena.

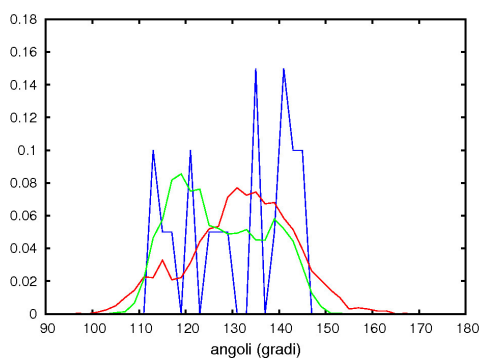
Figura I.2: Distribuzioni normalizzate degli angoli del linker. La linea rossa rappresenta i valori ottenuti dalla simulazione usante il modello bistabile implementato sia sugli angoli che sui diedri del linker, la linea verde i dati ricavati dalla simulazione con il potenziale quartico applicato agli angoli del linker, e la linea blu i dati delle strutture NMR.



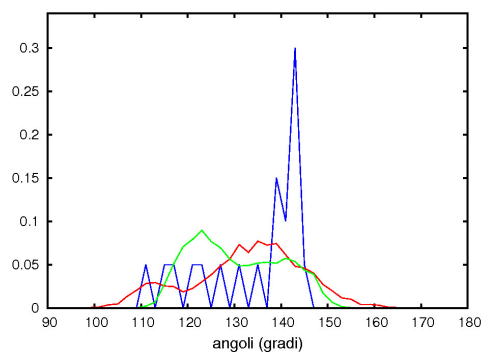
(a) Posizione 132 lungo la catena.



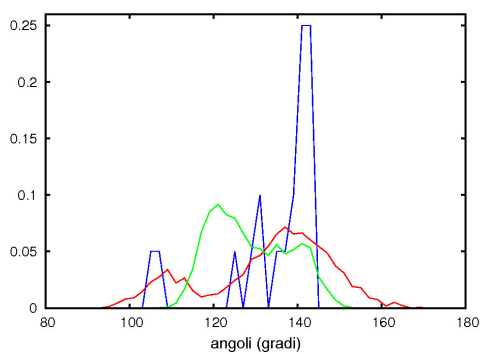
(b) Posizione 133 lungo la catena.



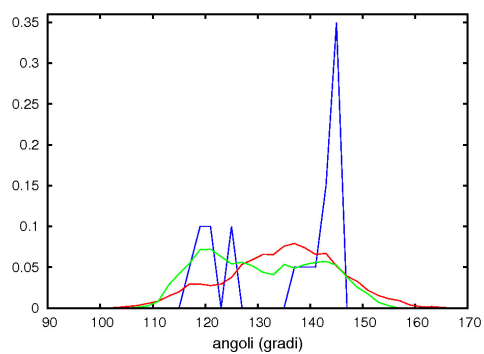
(c) Posizione 134 lungo la catena.



(d) Posizione 135 lungo la catena.

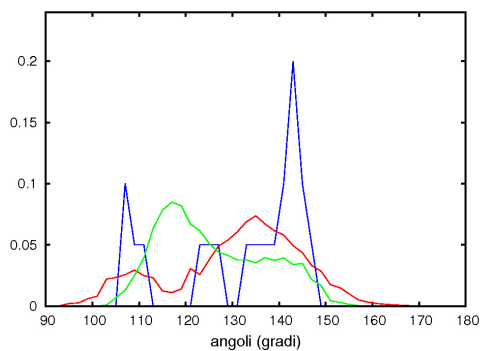


(e) Posizione 136 lungo la catena.

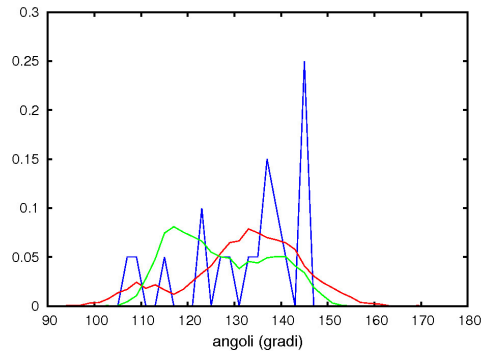


(f) Posizione 137 lungo la catena.

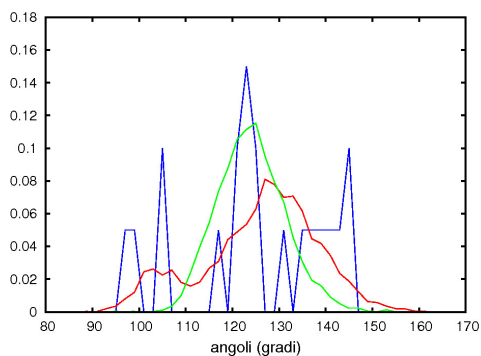
Figura I.3: Distribuzioni normalizzate degli angoli del linker. La linea rossa rappresenta i valori ottenuti dalla simulazione usando il modello bistabile implementato sia sugli angoli che sui diedri del linker, la linea verde i dati ricavati dalla simulazione con il potenziale quartico applicato agli angoli del linker, e la linea blu i dati delle strutture NMR.



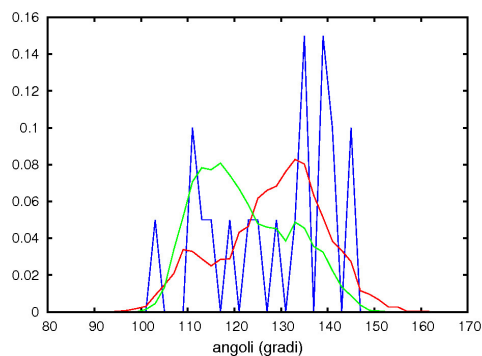
(a) Posizione 138 lungo la catena.



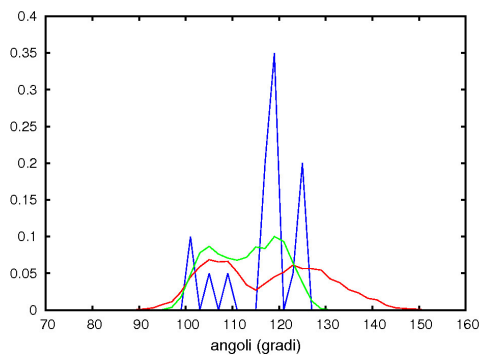
(b) Posizione 139 lungo la catena.



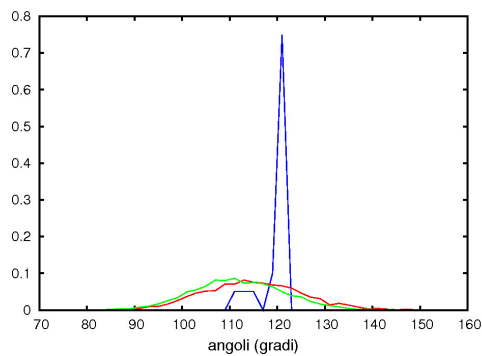
(c) Posizione 140 lungo la catena.



(d) Posizione 141 lungo la catena.

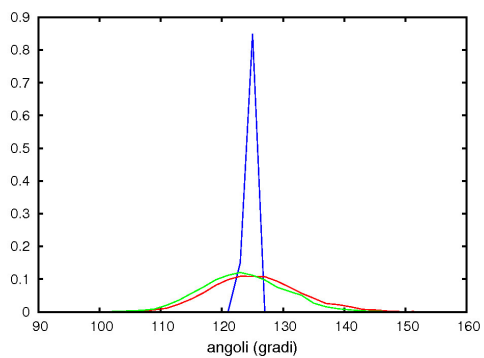


(e) Posizione 142 lungo la catena.

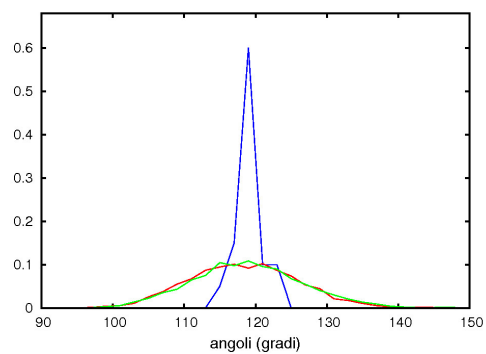


(f) Posizione 143 lungo la catena.

Figura I.4: Distribuzioni normalizzate degli angoli del linker. La linea rossa rappresenta i valori ottenuti dalla simulazione usando il modello bistabile implementato sia sugli angoli che sui diedri del linker, la linea verde i dati ricavati dalla simulazione con il potenziale quartico applicato agli angoli del linker, e la linea blu i dati delle strutture NMR.



(a) *Posizione 144 lungo la catena.*



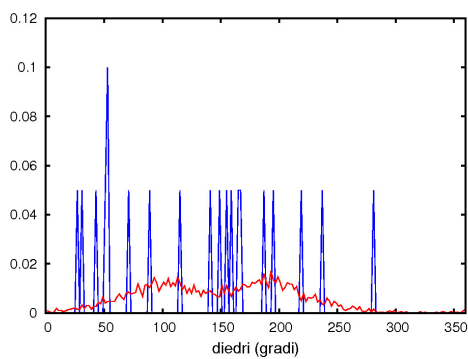
(b) *Posizione 145 lungo la catena.*

Figura I.5: Distribuzioni normalizzate degli angoli del linker. La linea rossa rappresenta i valori ottenuti dalla simulazione usante il modello bistabile implementato sia sugli angoli che sui diedri del linker, la linea verde i dati ricavati dalla simulazione con il potenziale quartico applicato agli angoli del linker, e la linea blu i dati delle strutture NMR.

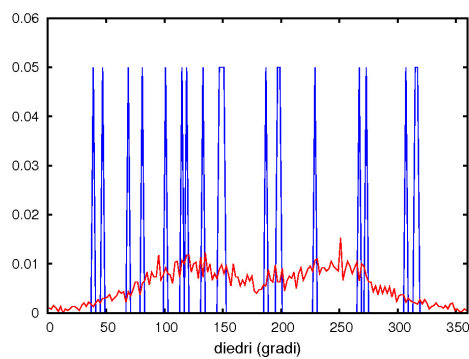
Appendice J

Distribuzioni dei diedri del linker

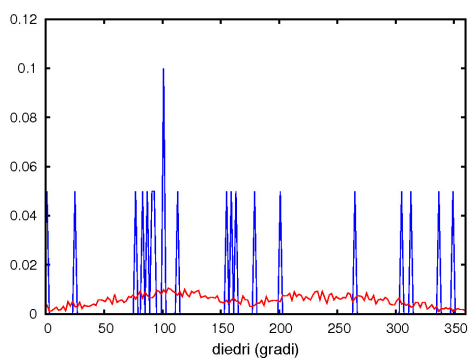
Il modello bistabile è stato applicato ai diedri del linker aventi le seguenti posizioni lungo la catena: 120, 121, 122, 123, 124, 125, 126, 127, 128, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140.



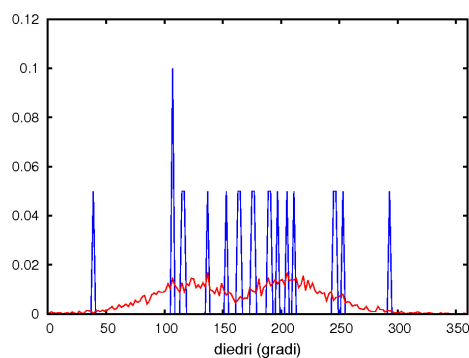
(a) Posizione 120 lungo la catena.



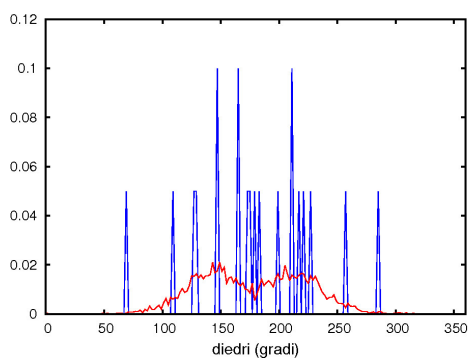
(b) Posizione 121 lungo la catena.



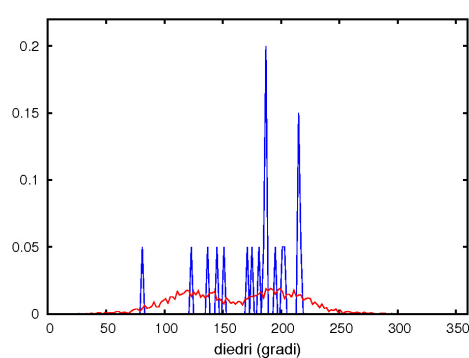
(c) Posizione 122 lungo la catena.



(d) Posizione 123 lungo la catena.

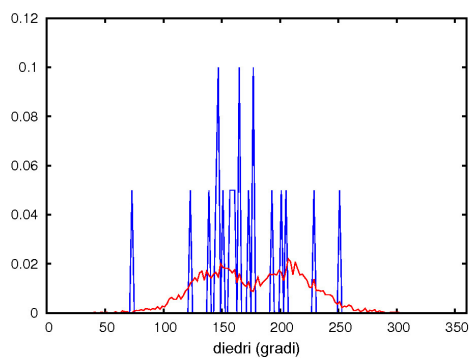


(e) Posizione 124 lungo la catena.

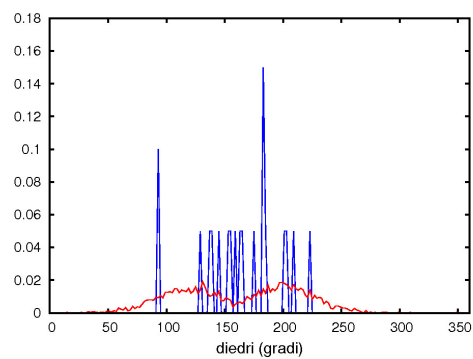


(f) Posizione 125 lungo la catena.

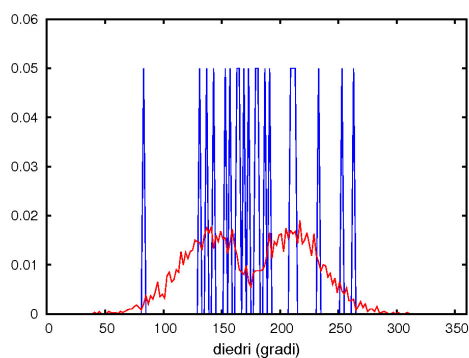
Figura J.1: Distribuzioni normalizzate dei diedri del linker. La linea rossa rappresenta i valori ottenuti dalla simulazione usante il modello bistabile implementato sia sugli angoli che sui diedri del linker e la linea blu i dati delle strutture NMR.



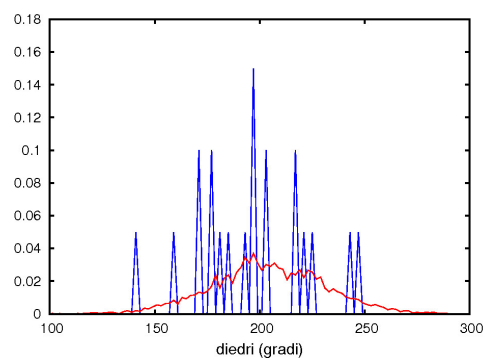
(a) *Posizione 126 lungo la catena.*



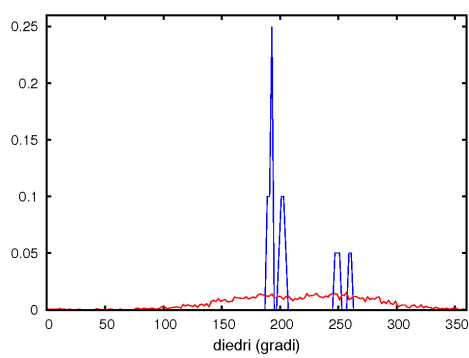
(b) *Posizione 127 lungo la catena.*



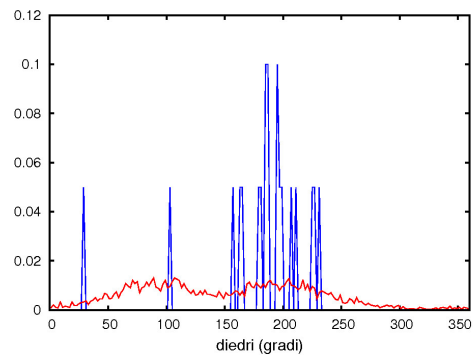
(c) *Posizione 128 lungo la catena.*



(d) *Posizione 129 lungo la catena.*

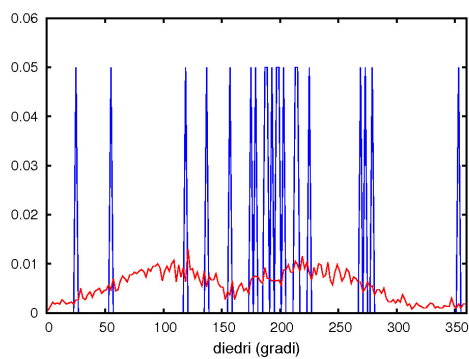


(e) *Posizione 130 lungo la catena.*

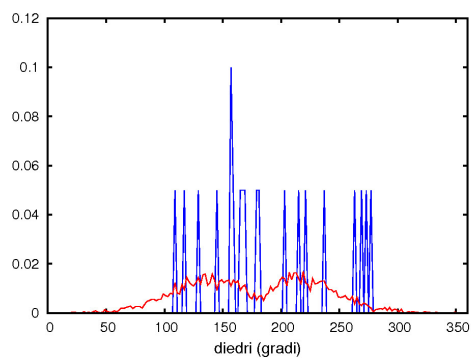


(f) *Posizione 131 lungo la catena.*

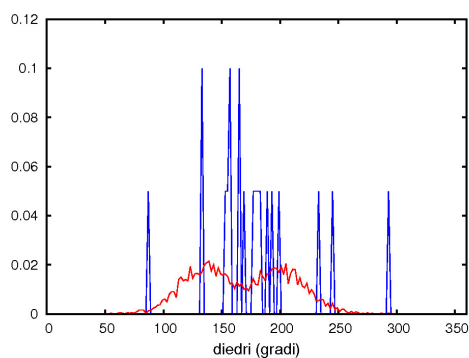
Figura J.2: Distribuzioni normalizzate dei diedri del linker. La linea rossa rappresenta i valori ottenuti dalla simulazione usante il modello bistabile implementato sia sugli angoli che sui diedri del linker e la linea blu i dati delle strutture NMR.



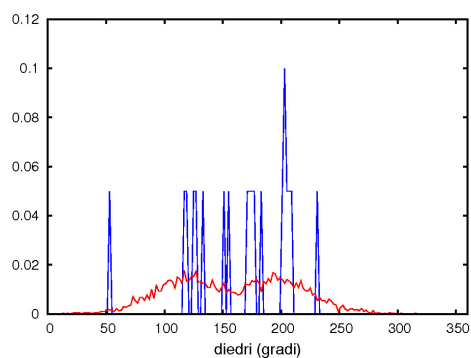
(a) Posizione 132 lungo la catena.



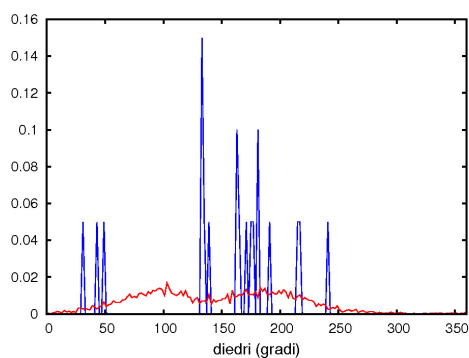
(b) Posizione 133 lungo la catena.



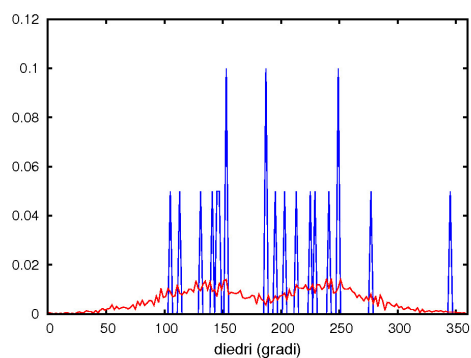
(c) Posizione 134 lungo la catena.



(d) Posizione 135 lungo la catena.

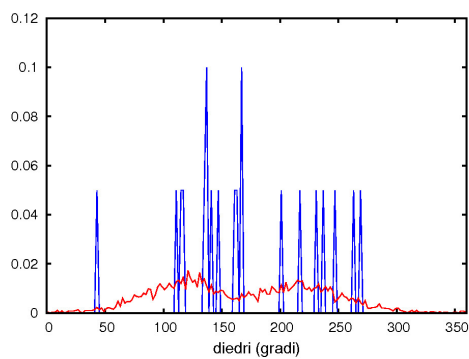


(e) Posizione 136 lungo la catena.

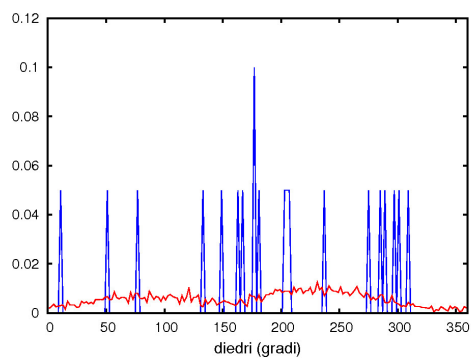


(f) Posizione 137 lungo la catena.

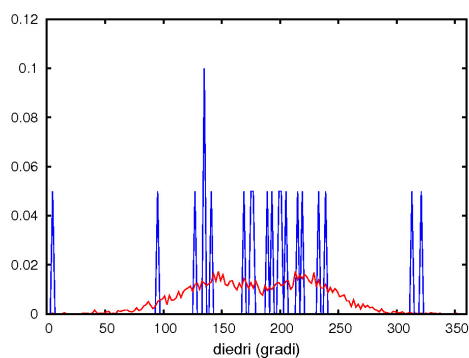
Figura J.3: Distribuzioni normalizzate dei diedri del linker. La linea rossa rappresenta i valori ottenuti dalla simulazione usante il modello bistabile implementato sia sugli angoli che sui diedri del linker e la linea blu i dati delle strutture NMR.



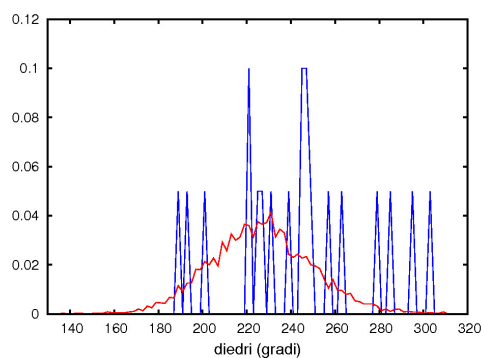
(a) Posizione 138 lungo la catena.



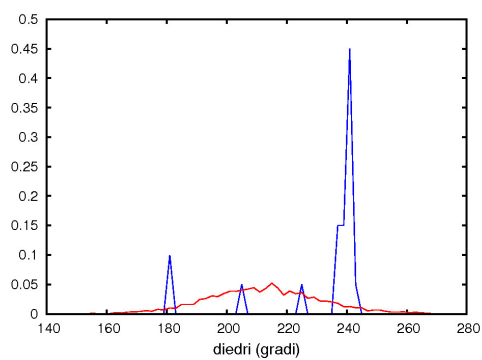
(b) Posizione 139 lungo la catena.



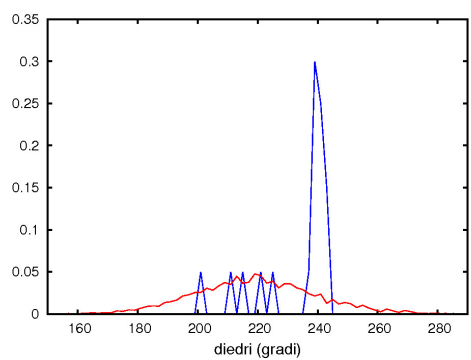
(c) Posizione 140 lungo la catena.



(d) Posizione 141 lungo la catena.

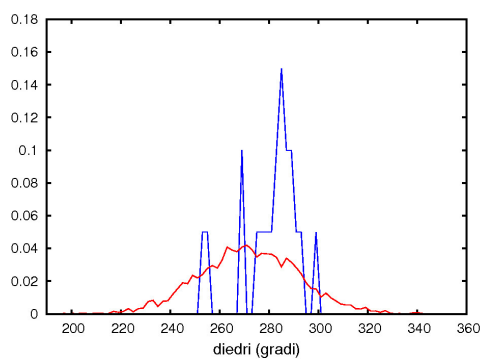


(e) Posizione 142 lungo la catena.

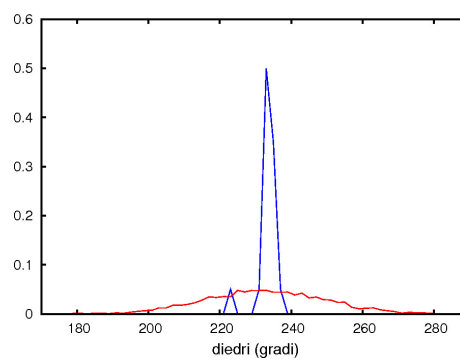


(f) Posizione 143 lungo la catena.

Figura J.4: Distribuzioni normalizzate dei diedri del linker. La linea rossa rappresenta i valori ottenuti dalla simulazione usante il modello bistabile implementato sia sugli angoli che sui diedri del linker e la linea blu i dati delle strutture NMR.



(a) *Posizione 144 lungo la catena.*



(b) *Posizione 145 lungo la catena.*

Figura J.5: Distribuzioni normalizzate dei diedri del linker. La linea rossa rappresenta i valori ottenuti dalla simulazione usante il modello bistabile implementato sia sugli angoli che sui diedri del linker e la linea blu i dati delle strutture NMR.

Appendice K

Grafici delle simulazioni dell'interazione Pr-Gag

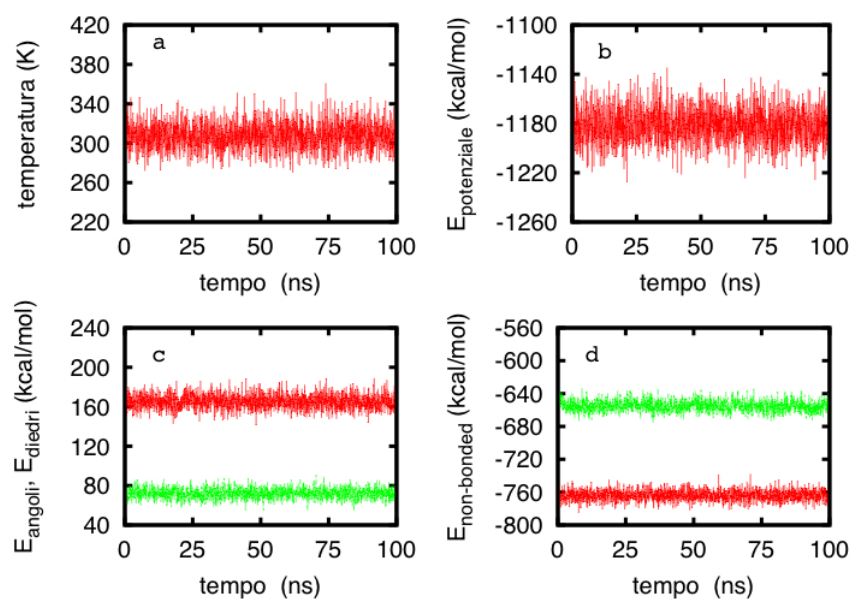


Figura K.1: Grafici in funzione del tempo della temperatura (a), dell'energia potenziale (b), del potenziale degli angoli (linea rossa in (c)) e dei diedri (linea verde in (c)), e del potenziale dato dai termini non bonded locali (linea rossa in (d)) e dai non bonded non locali (linea verde in (d)) riguardanti il complesso Pr-Gag, con la proteasi agganciata al sito di taglio, con il campo di forze n°1 e la struttura di partenza '2457'.

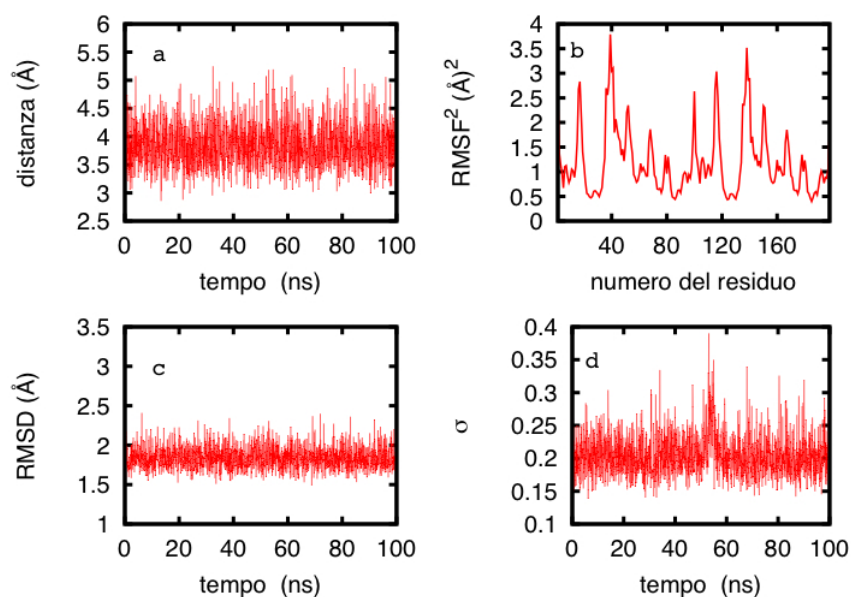


Figura K.2: Grafici della distanza tra le punte dei flap in funzione del tempo della simulazione della proteasi con la poliproteina Gag (a), di RMSF^2 (b), di RMSD (c), e della variabile σ al variare del tempo (d) riguardanti il complesso Pr-Gag, con la proteasi agganciata al sito di taglio, con il campo di forze n°1 e la struttura di partenza '2457'.

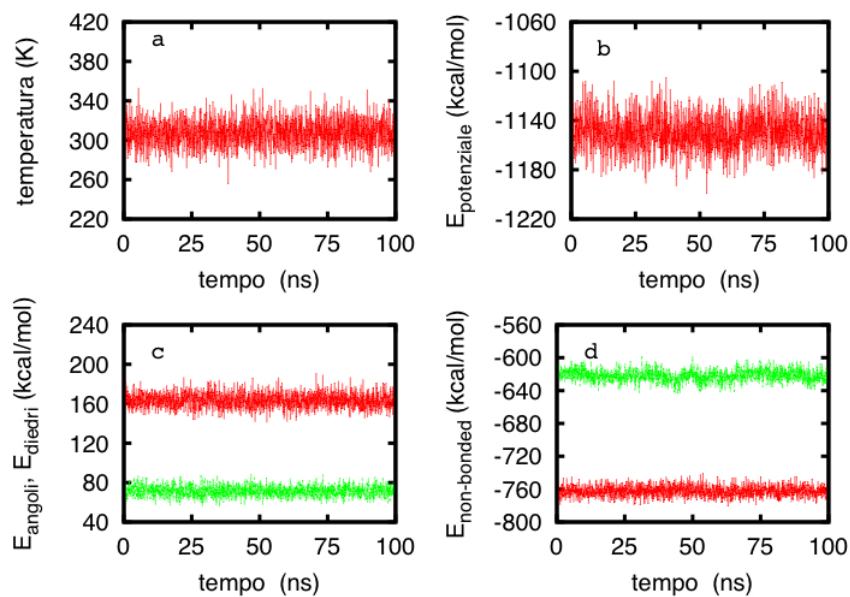


Figura K.3: Grafici in funzione del tempo della temperatura (a), dell'energia potenziale (b), del potenziale degli angoli (linea rossa in (c)) e dei diedri (linea verde in (c)), e del potenziale dato dai termini non bonded locali (linea rossa in (d)) e dai non bonded non locali (linea verde in (d)) riguardanti il complesso Pr-Gag, con la proteasi agganciata al sito di taglio, con il campo di forze n^o2 e la struttura di partenza '2457'.

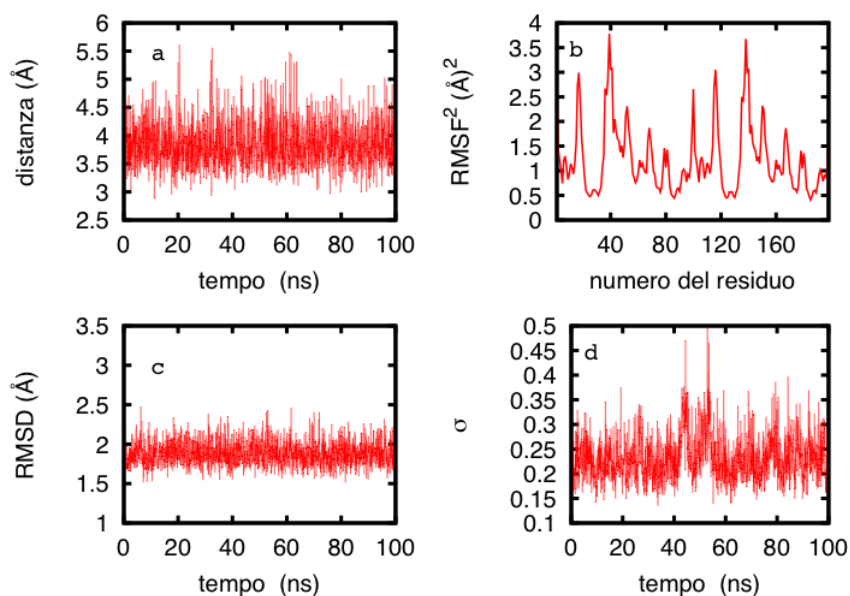


Figura K.4: Grafici della distanza tra le punte dei flap in funzione del tempo della simulazione della proteasi con la poliproteina Gag (a), di $RMSF^2$ (b), di RMSD (c), e della variabile σ al variare del tempo (d) riguardanti il complesso Pr-Gag, con la proteasi agganciata al sito di taglio, con il campo di forze n^o2 e la struttura di partenza '2457'.

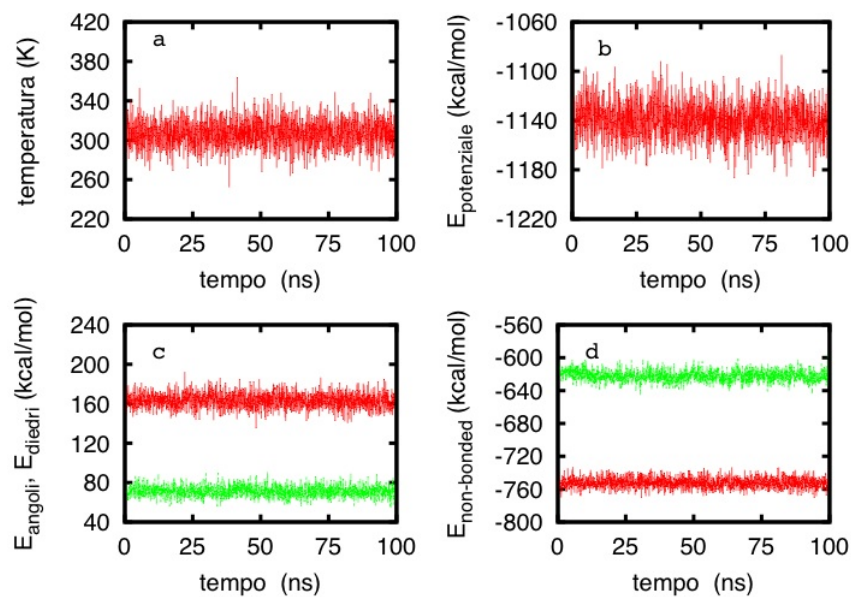


Figura K.5: Grafici in funzione del tempo della temperatura (a), dell'energia potenziale (b), del potenziale degli angoli (linea rossa in (c)) e dei diedri (linea verde in (c)), e del potenziale dato dai termini non bonded locali (linea rossa in (d)) e dai non bonded non locali (linea verde in (d)) riguardanti il complesso Pr-Gag, con la proteasi agganciata al sito di taglio, con il campo di forze n^o3 e la struttura di partenza '2457'.

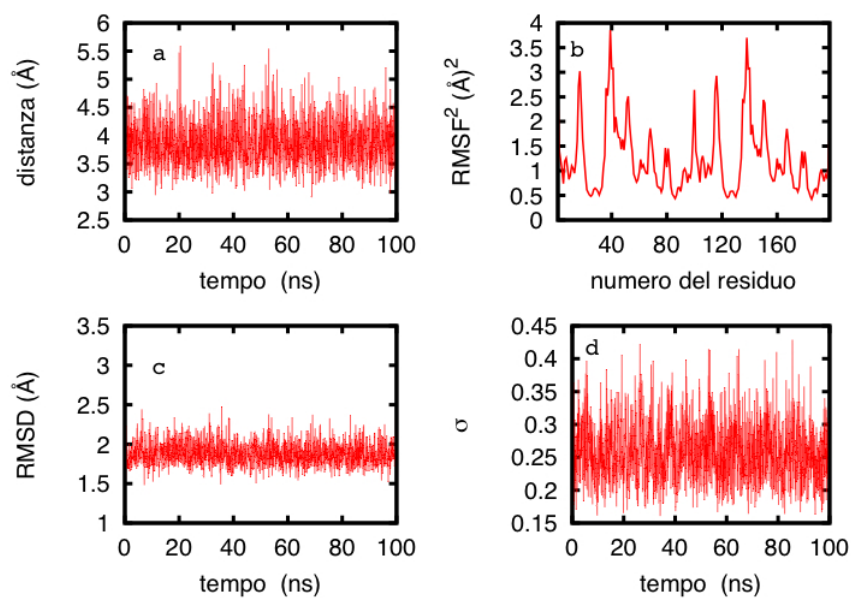


Figura K.6: Grafici della distanza tra le punte dei flap in funzione del tempo della simulazione della proteasi con la poliproteina Gag (a), di $RMSF^2$ (b), di RMSD (c), e della variabile σ al variare del tempo (d) riguardanti il complesso Pr-Gag, con la proteasi agganciata al sito di taglio, con il campo di forze n^o3 e la struttura di partenza '2457'.

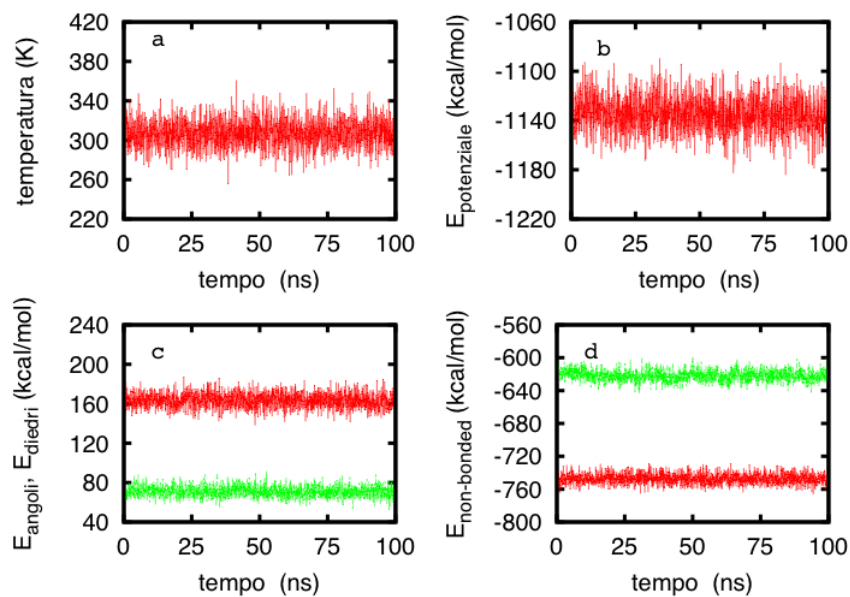


Figura K.7: Grafici in funzione del tempo della temperatura (a), dell'energia potenziale (b), del potenziale degli angoli (linea rossa in (c)) e dei diedri (linea verde in (c)), e del potenziale dato dai termini non bonded locali (linea rossa in (d)) e dai non bonded non locali (linea verde in (d)) riguardanti il complesso Pr-Gag, con la proteasi agganciata al sito di taglio, con il campo di forze n^o4 e la struttura di partenza '2457'.

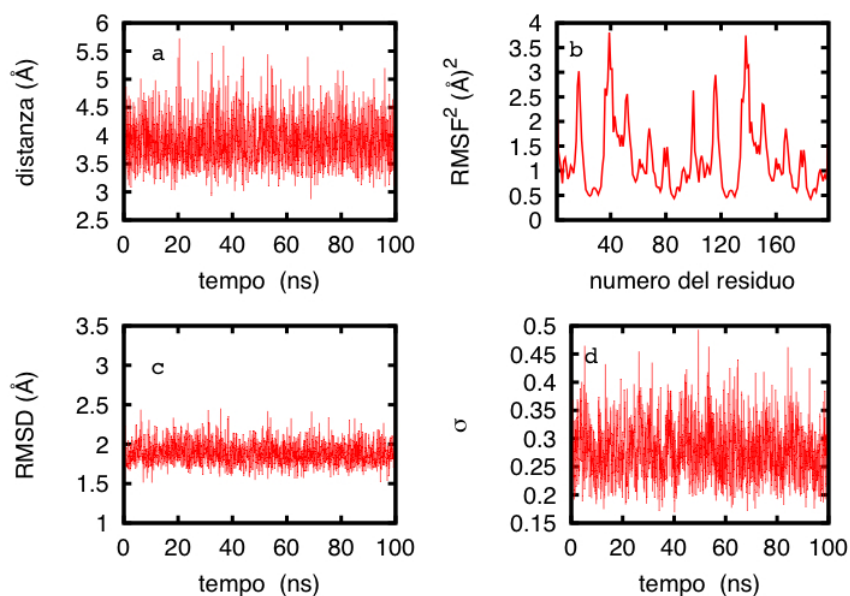


Figura K.8: Grafici della distanza tra le punte dei flap in funzione del tempo della simulazione della proteasi con la poliproteina Gag (a), di $RMSF^2$ (b), di RMSD (c), e della variabile σ al variare del tempo (d) riguardanti il complesso Pr-Gag, con la proteasi agganciata al sito di taglio, con il campo di forze n^o4 e la struttura di partenza '2457'.

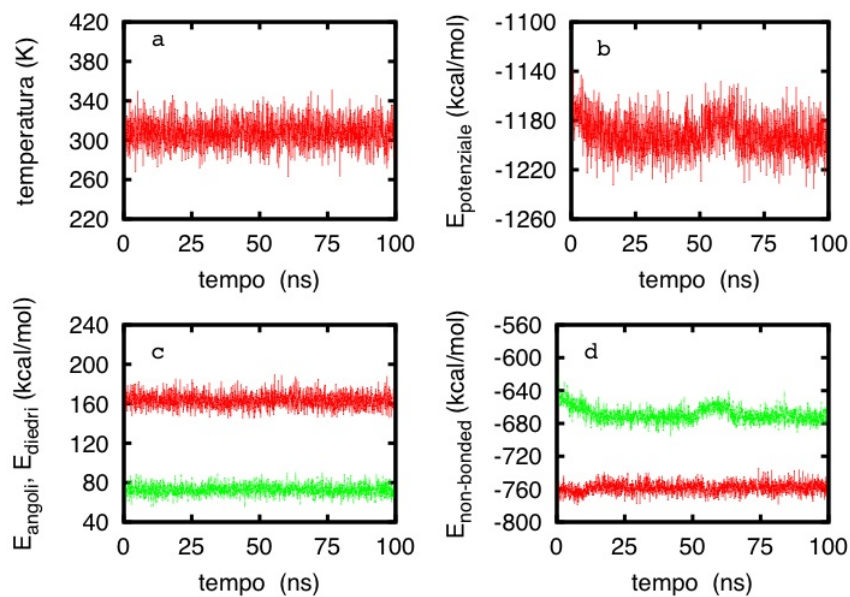


Figura K.9: Grafici in funzione del tempo della temperatura (a), dell'energia potenziale (b), del potenziale degli angoli (linea rossa in (c)) e dei diedri (linea verde in (c)), e del potenziale dato dai termini non bonded locali (linea rossa in (d)) e dai non bonded non locali (linea verde in (d)) riguardanti il complesso Pr-Gag, con la proteasi agganciata al sito di taglio, con il campo di forze n°1 e la struttura di partenza '3002'.

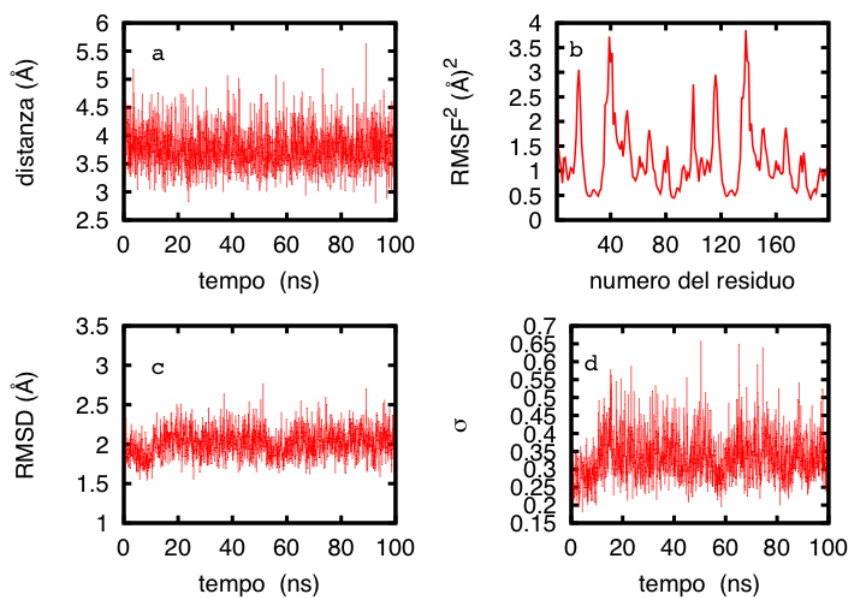


Figura K.10: Grafici della distanza tra le punte dei flap in funzione del tempo della simulazione della proteasi con la poliproteina Gag (a), di $RMSF^2$ (b), di RMSD (c), e della variabile σ al variare del tempo (d) riguardanti il complesso Pr-Gag, con la proteasi agganciata al sito di taglio, con il campo di forze n°1 e la struttura di partenza '3002'.

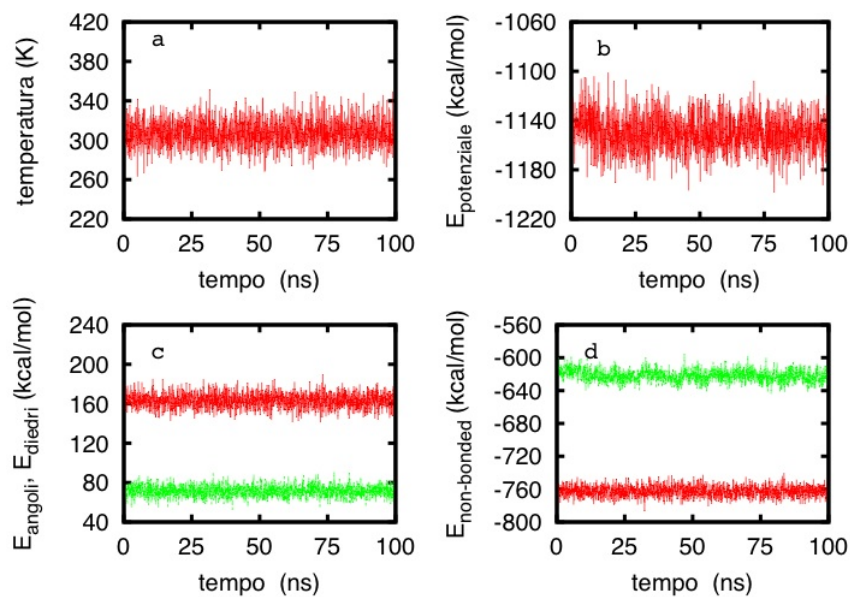


Figura K.11: Grafici in funzione del tempo della temperatura (a), dell'energia potenziale (b), del potenziale degli angoli (linea rossa in (c)) e dei diedri (linea verde in (c)), e del potenziale dato dai termini non bonded locali (linea rossa in (d)) e dai non bonded non locali (linea verde in (d)) riguardanti il complesso Pr-Gag, con la proteasi agganciata al sito di taglio, con il campo di forze n^o2 e la struttura di partenza '3002'.

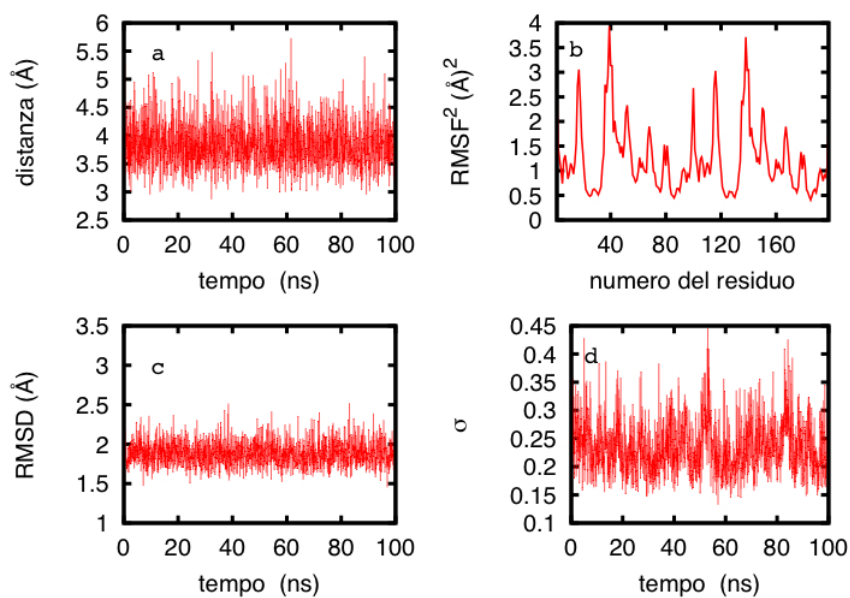


Figura K.12: Grafici della distanza tra le punte dei flap in funzione del tempo della simulazione della proteasi con la poliproteina Gag (a), di $RMSF^2$ (b), di RMSD (c), e della variabile σ al variare del tempo (d) riguardanti il complesso Pr-Gag, con la proteasi agganciata al sito di taglio, con il campo di forze n^o2 e la struttura di partenza '3002'.

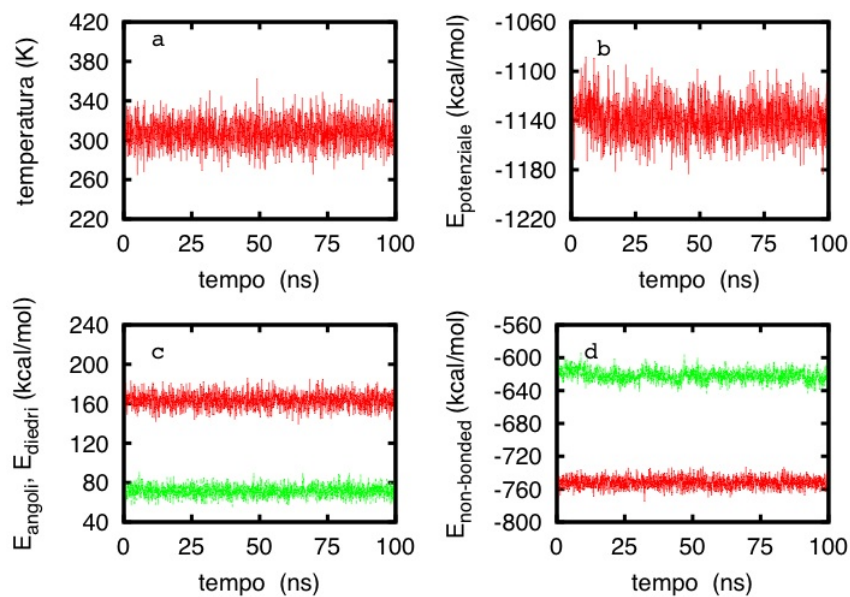


Figura K.13: Grafici in funzione del tempo della temperatura (a), dell'energia potenziale (b), del potenziale degli angoli (linea rossa in (c)) e dei diedri (linea verde in (c)), e del potenziale dato dai termini non bonded locali (linea rossa in (d)) e dai non bonded non locali (linea verde in (d)) riguardanti il complesso Pr-Gag, con la proteasi agganciata al sito di taglio, con il campo di forze $n^{\circ}3$ e la struttura di partenza '3002'.

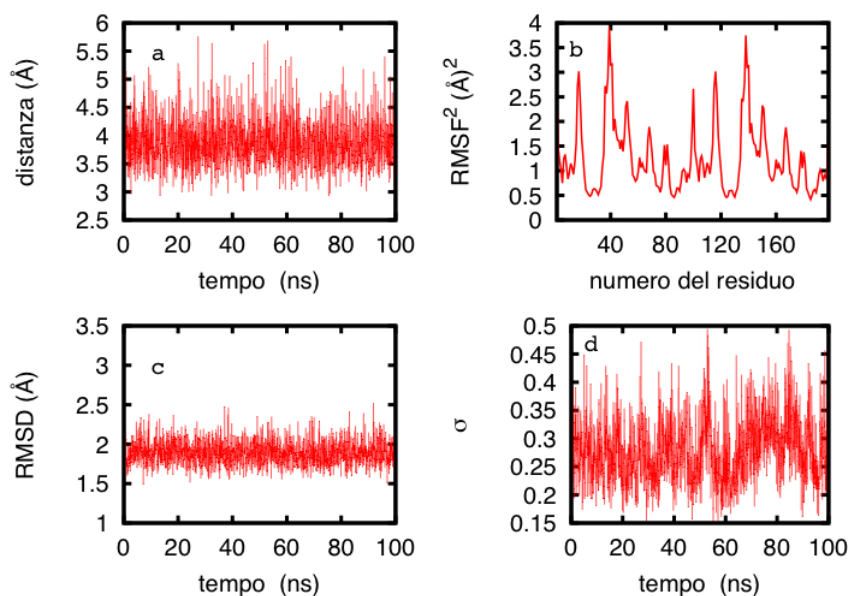


Figura K.14: Grafici della distanza tra le punte dei flap in funzione del tempo della simulazione della proteasi con la poliproteina Gag (a), di $RMSF^2$ (b), di RMSD (c), e della variabile σ al variare del tempo (d) riguardanti il complesso Pr-Gag, con la proteasi agganciata al sito di taglio, con il campo di forze $n^{\circ}3$ e la struttura di partenza '3002'.

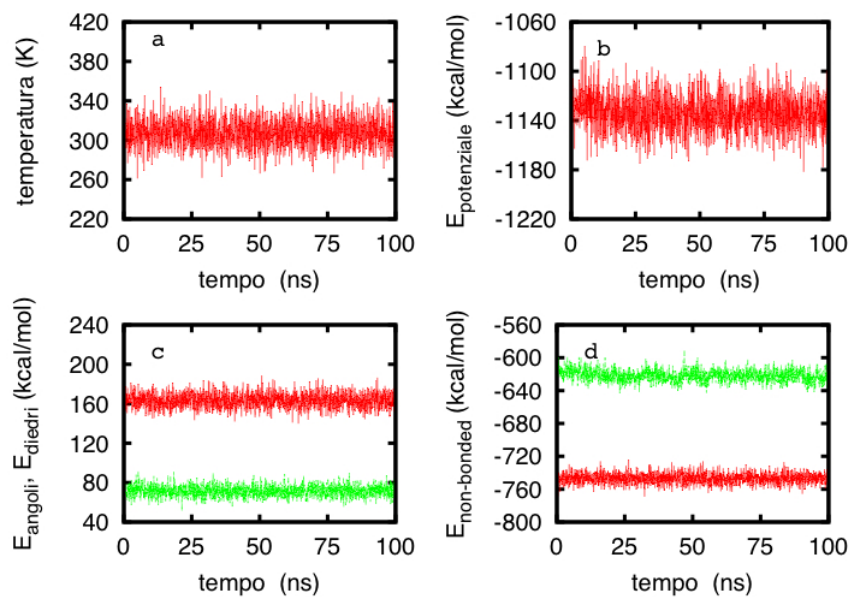


Figura K.15: Grafici in funzione del tempo della temperatura (a), dell'energia potenziale (b), del potenziale degli angoli (linea rossa in (c)) e dei diedri (linea verde in (c)), e del potenziale dato dai termini non bonded locali (linea rossa in (d)) e dai non bonded non locali (linea verde in (d)) riguardanti il complesso Pr-Gag, con la proteasi agganciata al sito di taglio, con il campo di forze n^o4 e la struttura di partenza '3002'.

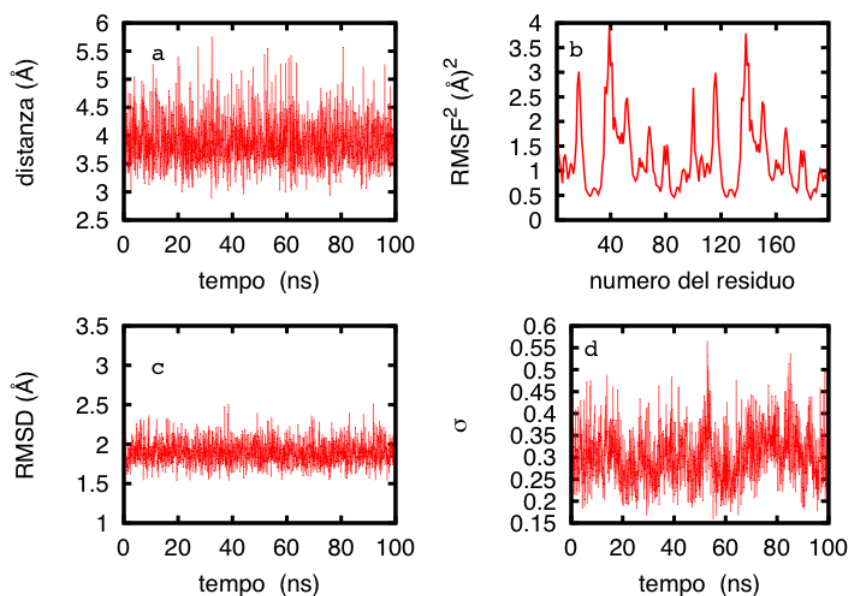


Figura K.16: Grafici della distanza tra le punte dei flap in funzione del tempo della simulazione della proteasi con la poliproteina Gag (a), di $RMSF^2$ (b), di RMSD (c), e della variabile σ al variare del tempo (d) riguardanti il complesso Pr-Gag, con la proteasi agganciata al sito di taglio, con il campo di forze n^o4 e la struttura di partenza '3002'.

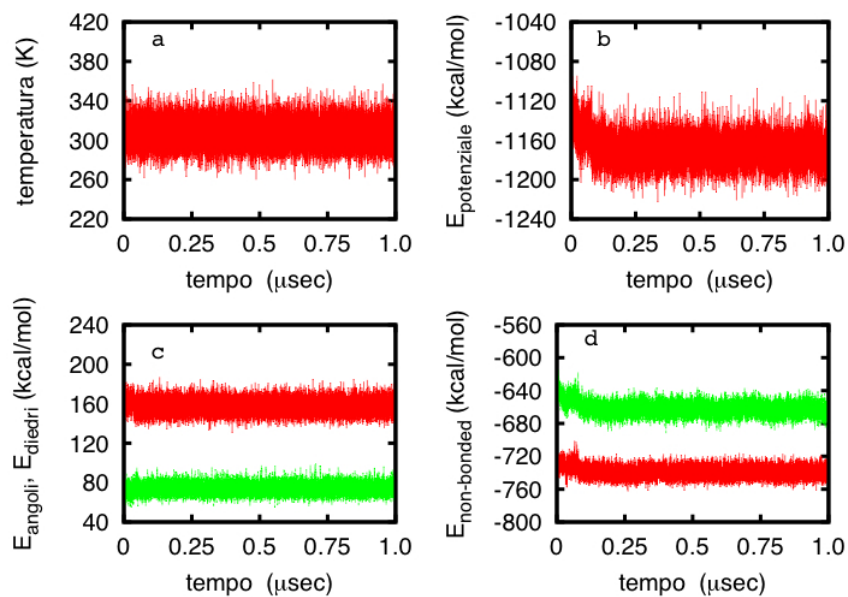


Figura K.17: Grafici in funzione del tempo della temperatura (a), dell'energia potenziale (b), del potenziale degli angoli (linea rossa in (c)) e dei diedri (linea verde in (c)), e del potenziale dato dai termini non bonded locali (linea rossa in (d)) e dai non bonded non locali (linea verde in (d)) riguardanti Gag e Proteasi inizialmente separate, con il campo di forze n°1 e la struttura di partenza '2457'.

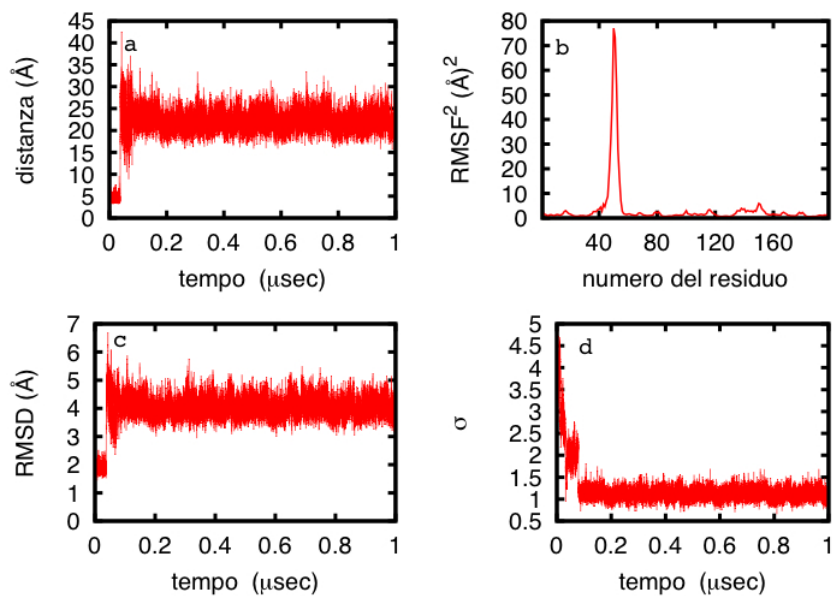


Figura K.18: Grafici della distanza tra le punte dei flap in funzione del tempo della simulazione della proteasi con la poliproteina Gag (a), di $RMSF^2$ (b), di RMSD (c), e della variabile σ al variare del tempo (d) riguardanti Gag e Proteasi inizialmente separate, con il campo di forze n°1 e la struttura di partenza '2457'.

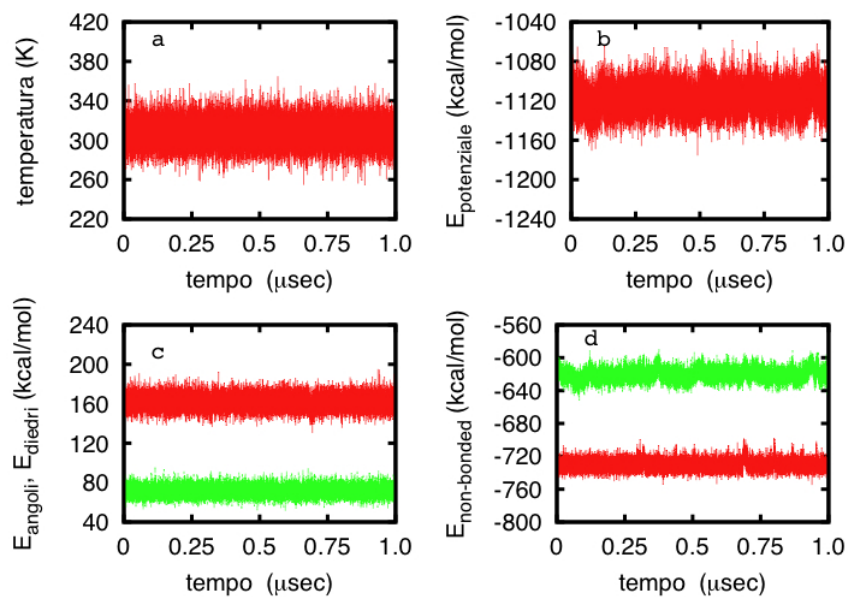


Figura K.19: Grafici in funzione del tempo della temperatura (a), dell'energia potenziale (b), del potenziale degli angoli (linea rossa in (c)) e dei diedri (linea verde in (c)), e del potenziale dato dai termini non bonded locali (linea rossa in (d)) e dai non bonded non locali (linea verde in (d)) riguardanti Gag e Proteasi inizialmente separate, con il campo di forze n^o2 e la struttura di partenza '2457'.

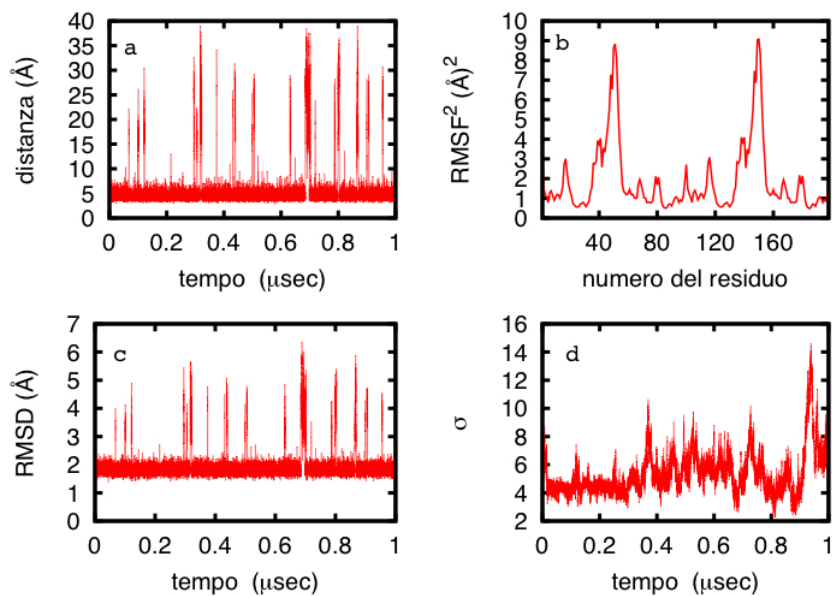


Figura K.20: Grafici della distanza tra le punte dei flap in funzione del tempo della simulazione della proteasi con la poliproteina Gag (a), di $RMSF^2$ (b), di RMSD (c), e della variabile σ al variare del tempo (d) riguardanti Gag e Proteasi inizialmente separate, con il campo di forze n^o2 e la struttura di partenza '2457'.

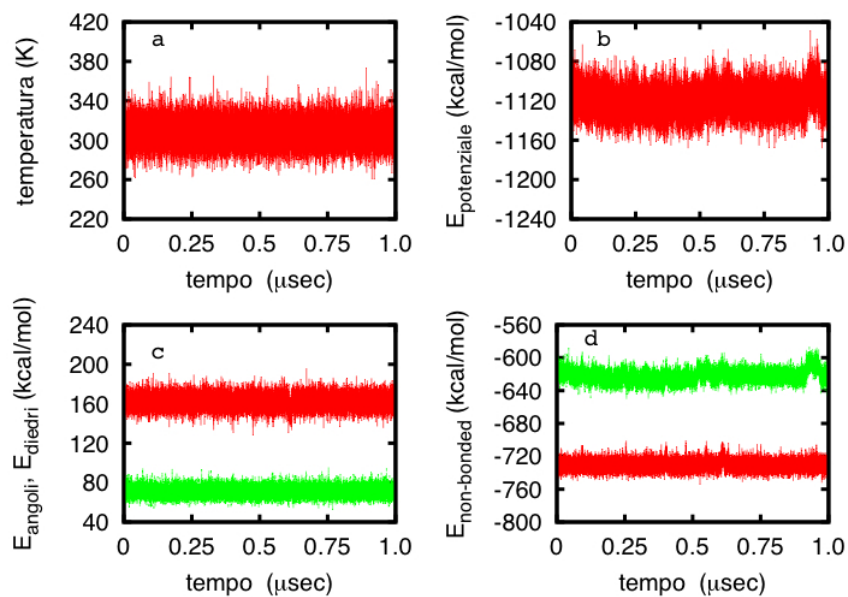


Figura K.21: Grafici in funzione del tempo della temperatura (a), dell'energia potenziale (b), del potenziale degli angoli (linea rossa in (c)) e dei diedri (linea verde in (c)), e del potenziale dato dai termini non bonded locali (linea rossa in (d)) e dai non bonded non locali (linea verde in (d)) riguardanti Gag e Proteasi inizialmente separate, con il campo di forze n^o4 e la struttura di partenza '2457'.

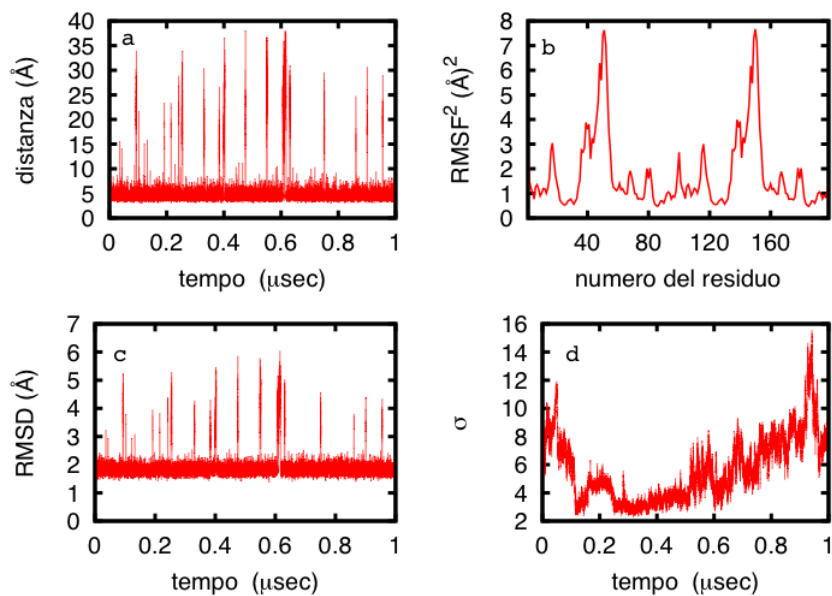


Figura K.22: Grafici della distanza tra le punte dei flap in funzione del tempo della simulazione della proteasi con la poliproteina Gag (a), di $RMSF^2$ (b), di RMSD (c), e della variabile σ al variare del tempo (d) riguardanti Gag e Proteasi inizialmente separate, con il campo di forze n^o4 e la struttura di partenza '2457'.

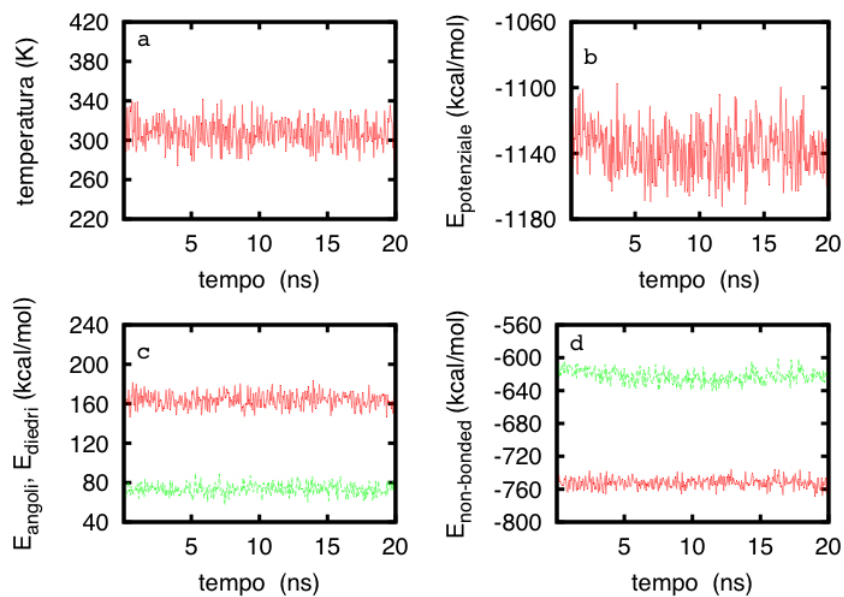


Figura K.23: Grafici in funzione del tempo della temperatura (a), dell'energia potenziale (b), del potenziale degli angoli (linea rossa in (c)) e dei diedri (linea verde in (c)), e del potenziale dato dai termini non bonded locali (linea rossa in (d)) e dai non bonded non locali (linea verde in (d)) della simulazione in cui pr è spostata inizialmente di 2 amminoacidi rispetto al sito di taglio, con il campo di forze n°3.

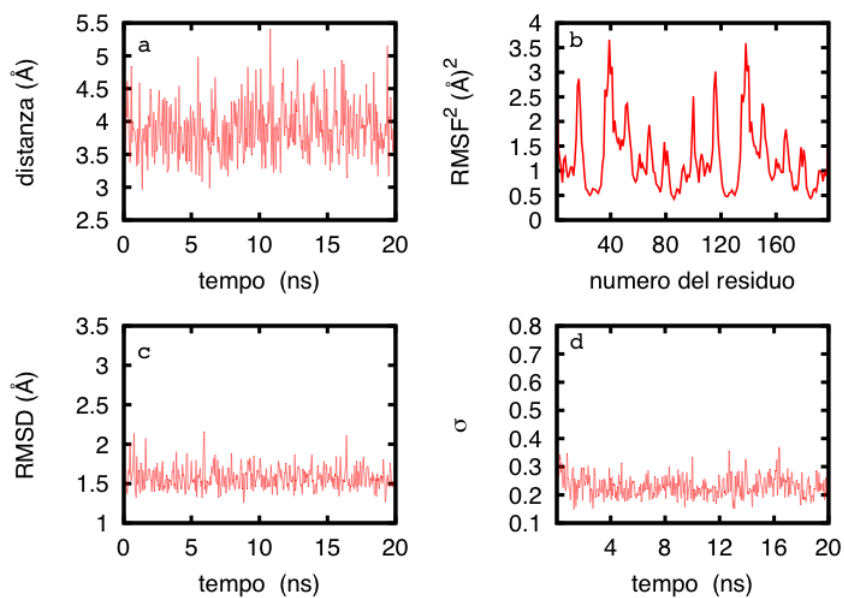


Figura K.24: Grafici della distanza tra le punte dei flap in funzione del tempo della simulazione della proteasi con la poliproteina Gag (a), di $RMSF^2$ (b), di RMSD (c), e della variabile σ al variare del tempo (d) della simulazione in cui pr è spostata inizialmente di 2 amminoacidi rispetto al sito di taglio, con il campo di forze n°3.

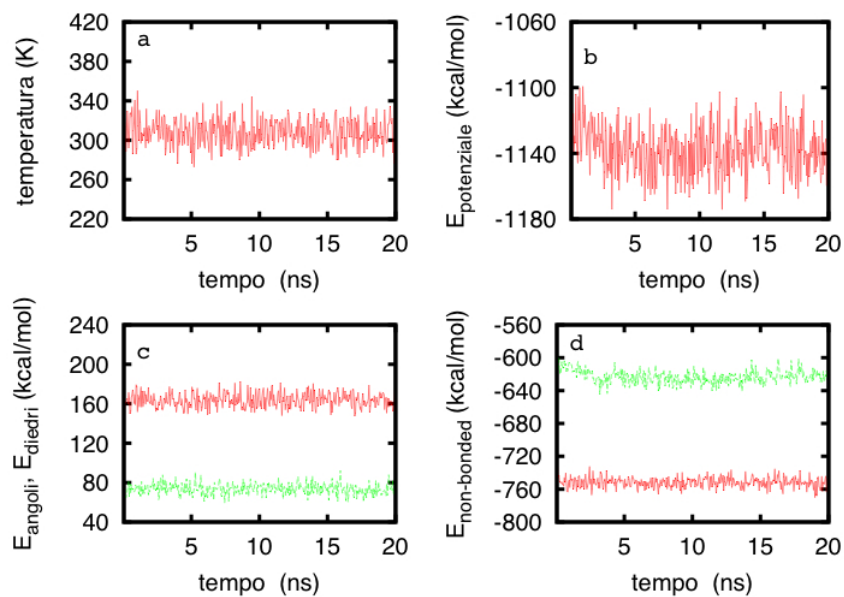


Figura K.25: Grafici in funzione del tempo della temperatura (a), dell'energia potenziale (b), del potenziale degli angoli (linea rossa in (c)) e dei diedri (linea verde in (c)), e del potenziale dato dai termini non bonded locali (linea rossa in (d)) e dai non bonded non locali (linea verde in (d)) della simulazione in cui pr è spostata inizialmente di 4 amminoacidi rispetto al sito di taglio, con il campo di forze n°3.

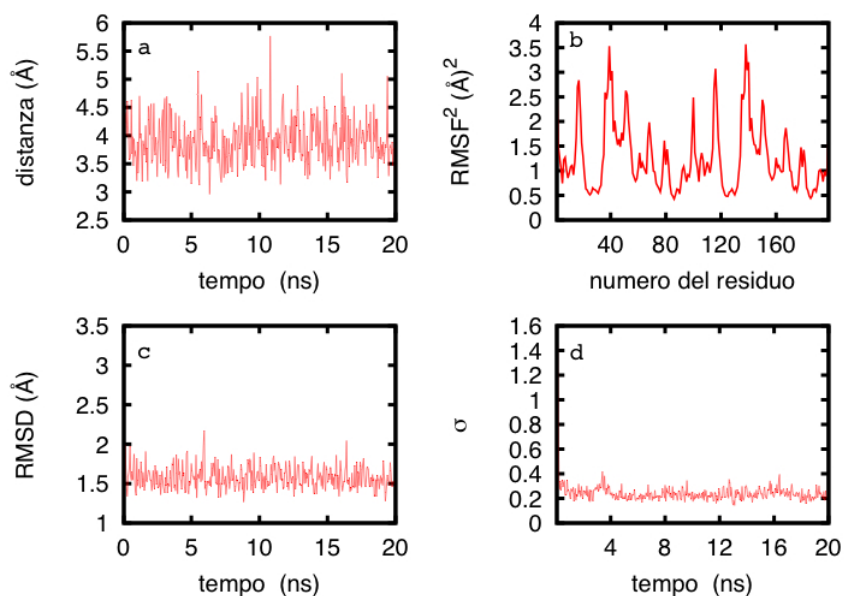


Figura K.26: Grafici della distanza tra le punte dei flap in funzione del tempo della simulazione della proteasi con la poliproteina Gag (a), di $RMSF^2$ (b), di RMSD (c), e della variabile σ al variare del tempo (d) della simulazione in cui pr è spostata inizialmente di 4 amminoacidi rispetto al sito di taglio, con il campo di forze n°3.

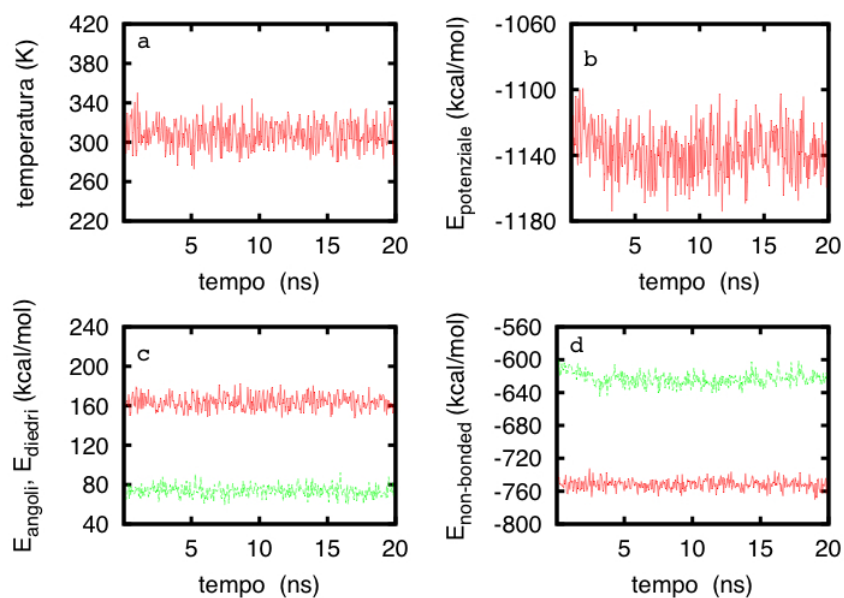


Figura K.27: Grafici in funzione del tempo della temperatura (a), dell'energia potenziale (b), del potenziale degli angoli (linea rossa in (c)) e dei diedri (linea verde in (c)), e del potenziale dato dai termini non bonded locali (linea rossa in (d)) e dai non bonded non locali (linea verde in (d)) della simulazione in cui pr è spostata inizialmente di 4 amminoacidi rispetto al sito di taglio, con la struttura di partenza di Gag che è stata ricavata dopo un riassetamento dell'intero sistema Gag-pr. Il campo di forze usato è il n^o3.

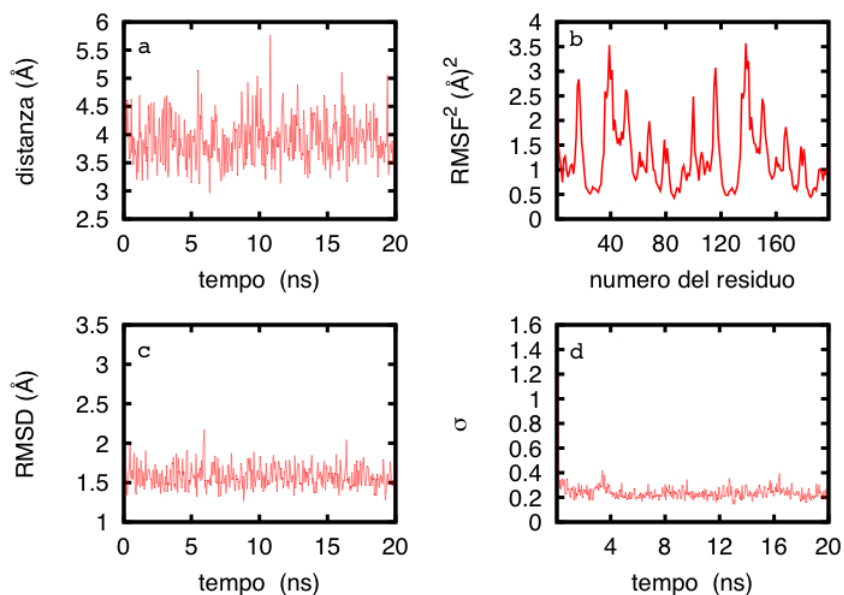


Figura K.28: Grafici della distanza tra le punte dei flap in funzione del tempo della simulazione della proteasi con la poliproteina Gag (a), di $RMSF^2$ (b), di RMSD (c), e della variabile σ al variare del tempo (d) della simulazione in cui pr è spostata inizialmente di 4 amminoacidi rispetto al sito di taglio, con la struttura di partenza di Gag che è stata ricavata dopo un riassetamento dell'intero sistema Gag-pr. Il campo di forze usato è il n^o3.

Bibliografia

- [1] Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, et al (1983) *Science* **220**, 868-871
- [2] Gallo, R. C., Salahuddin, S. Z., Popovic, M., Shearer, G. M., Kaplan, M., Haynes, B. F., Palker, T. J., Redfield, R., Oleske, J., Safai, B., White, G., Foster, P., and Markham, P. D. (1984) *Science* **224**, 500-503
- [3] Rousseau CM, Nduati RW, Richardson BA, John-Stewart GC, Mbori-Ngacha DA, Kreiss JK, Overbaugh J. (2004) *J Infect Dis.* **190(10)**, 1880-1888
- [4] Paul M. Sharp and Beatrice H. Hahn (2011) *Cold Spring Harb Perspect Med* **1**:a006841
- [5] Baltimore D. (1971) *Bacteriological Reviews* **35**, 235-241
- [6] Coffin JM, Hughes SH, Varmus HE (1997) *Retroviruses* Cold Spring Harbor Laboratory Press
- [7] Jacks, Power, Masiarz, Luciw, Barr, and Varmus, (1988) *Nature* **331**, 280-283
- [8] Ganser-Pornillos, Yeager, Sundquist (2008) *Current Opinion in Structural Biology* **18**, 203-217
- [9] Worthylake et al. (1999) *Acta Crystallography* **D55**, 85-92
- [10] Sundquist and Kräusslich (2012) *Cold Spring Harb Perspect Med* **2**:a006924
- [11] Fuller et al. (2001) *Journal of virology* **75**, 759-771
- [12] Wesley I Sundquist, Christopher P. Hill et al. (2006) *Biochemistry* **45**, 11257-11266
- [13] Barbie K Ganser-Pornillos, Mark Yeager and Wesley I Sundquist (2008) *Current Opinion in Structural Biology* **18**, 203-217
- [14] Briggs e Kräusslich (2011) *Journal of Molecular Biology* **410**, 491-500
- [15] Yeager and Sundquist (2008) *Current Opinion in Structural Biology* **18**, 203-217
- [16] Erik De Clercq (2004) *Nature Reviews Microbiology* **2**, 704-720
- [17] Kohl, N.E.; Emini, E.A.; Schleif, W.A.; Davis, L.J.; Heimbach, J.C.; Dixon, R.A.; Scolnick, E.M.; Sigal, I.S. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 4686-4690

- [18] Tomasselli A, Heinrikson R. (2000) *Biochim Biophys Acta* **1477**, 189-214
- [19] Yi Mao (2011) *BMC Structural Biology* **11**:31
- [20] Eric Lefebvre and Celia A. Schiffer (2008) *AIDS Rev.* **10(3)**, 131-142
- [21] Kuo-Chen Chou (1996) *Analytical Biochemistry* **233**, 1-14
- [22] Suresh Karthik and Sanjib Senapati (2011) *Proteins* **79**, 1830-1840
- [23] Chia-En A. Chang, Joanna Trylska, Valentina Tozzini and J. Andrew McCammon (2007) *Chemical Biology and Drug Design* **69**, 5-13
- [24] Fabio Pietrucci, Fabrizio Marinelli, Paolo Carloni, and Alessandro Laio (2009) *Journal of the American Chemical Society* **131**, 11811-11818
- [25] Gergely Toth, and Attila Borics (2006) *Biochemistry* **45**, 6606-6614
- [26] Joanna Trylska, Valentina Tozzini, Chia-en A. Chang, and J. Andrew McCammon (2007) *Biophysical Journal* **92**, 4179-4187
- [27] Nan-jie Deng, Weihua Zheng, Emillio Gallicchio, and Ronald M. Levy (2011) *Journal of the American Chemical Society* **133**, 9387-9394
- [28] Anthony J.F. Griffiths, Susan R. Wessler, Richard C. Lewontin, William M. Gelbart, David T. Suzuki, Jeffrey H. Miller (2006) *Genetica. Principi di analisi formale Zanichelli*
- [29] Akbar Ali, Rajintha M. Bandaranayake, Yufeng Cai, Nancy M. King, Madhavi Kolli, Seema Mittal, Jennifer F. Murzycki, Madhavi N.L. Nalam, Ellen A. Nalivaika, Aysegül Özen, Moses M. Prabu-Jeyabalan, Kelly Thayer and Celia A. Schiffer (2010) *Viruses*, **2**, 2509-2535
- [30] Thomas D. Wu, Celia A. Schiffer, Matthew J. Gonzales, Jonathan Taylor, Rami Kantor, Sunwen Chou, Dennis Israelski, Andrew R. Zolopa, W. Jeffrey Fessel, and Robert W. Shafer (2003) *Journal of Virology*, **77**, 4836-4847
- [31] Soo-Yon Rhee, Jonathan Taylor, W. Jeffrey Fessel, David Kaufman, William Towner, Paolo Troia, Peter Ruane James Hellinger, Vivian Shirvani, Andrew Zolopa, and Robert W. Shafer (2010) *Antimicrobial Agents and Chemotherapy*, **54**, 4253-4261
- [32] Jeremy M Berg, John L Tymoczko, and Lubert Stryer (2002) *Biochemistry* W. H. Freeman
- [33] Liwen You, Daniel Garwicz, and Thorsteinn Rognvaldsson (2005) *Journal of Virology*, **79**, 12477-12486
- [34] Josef Vagner, Hongchang Qu, and Victor J. Hruby (2008) *Curr Opin Chem Biol.*, **12(3)**, 292-296
- [35] Thorsteinn Rognvaldsson and Liwen You (2004) *BMC Bioinformatics* **20**, 1702-1709

- [36] Liwen You (2006) *Computational Systems Bioinformatics Conference* **5**, 249-256
- [37] Bing Niu, Lin Lu, Liang Liu, Tian Hong Gu, Kai-Yan Feng, Wen-Cong Lu, Yu-Dong Cai (2009) *Journal of Computational Chemistry* **30**, 33-39
- [38] Thorsteinn Rögnvaldsson, Terence A Etchells, Liwen You, Daniel Garwicz, Ian Jarman and Paulo JG Lisboa (2009) *BMC Bioinformatics* **10**, 149-166
- [39] Hasan Ogul (2009) *BioSystems* **96**, 246-250
- [40] Gilhan Kim, Yeonjoo Kim, Heuseok Lim, Hyeoncheol Kim (2010) *Artificial Intelligence in Medicine* **48**, 83-89
- [41] Loris Nanni, Alessandra Lumini (2010) *Artificial Intelligence in Medicine* **48**, 43-50
- [42] Aysegül Özen, Türkan Haliloglu and Celia A. Schiffer (2011) *Journal of Molecular Biology* **410**, 726-744
- [43] Madhavi N. L. Nalam, Akbar Ali, Michael D. Altman, G .S. Kiran Kumar Reddy, Sripriya Chellappan, Visvaldas Kairys, Aysegul Ozen, Hong Cao, Michael K. Gilson, Bruce Tidor, Tariq M. Rana, and Celia A. Schiffer (2010) *Journal of Virology* **84**, 5368-5378
- [44] Michael D. Altman, Akbar Ali, G. S. Kiran Kumar Reddy, Madhavi N. L. Nalam, Saima Ghafoor Anjum, Hong Cao, Sripriya Chellappan, Visvaldas Kairys, Miguel X. Fernandes, Michael K. Gilson, Celia A. Schiffer, Tariq M. Rana, and Bruce Tidor (2008) *Journal of the American Chemical Society* **130**, 6099-6113
- [45] Becker et al. *Computational Biochemistry and Biophysics* (2001) Marcel Dekker, Inc.
- [46] Valentina Tozzini *Metodi di Modellizzazione Multi-scala*
- [47] W. Smith, T.R. Forester and I.T. Todorov (2009) *THE DL POLY 2 USER MANUAL* STFC Daresbury Laboratory
- [48] Victor Ruhle (2007) *Berendsen and Nose-Hoover thermostats*
- [49] Fabio Trovato and Valentina Tozzini (2012) *AIP Conf. Proc.* **187**, 1456
- [50] Konrad Hinsen (1998) *Proteins: Structure, Function, and Genetics* **33**, 417-429
- [51] Turkan Haliloglu, Ivet Bahar, and Burak Erman (1997) *Physical Review Letters* **79**, 417-429
- [52] Wenjun Zheng, Bernard R. Brooks, and Gerhard Hummer (2007) *Proteins: Structure, Function, and Genetics* **69**, 43-57
- [53] Jhih-Wei Chu and Gregory A. Voth (2007) *Biophysical Journal* **93**, 3860-3871
- [54] Nobuhiro GO and Harold A. Scheraga (1976) *Macromolecules* **9**, 535-542

- [55] Valentina Tozzini (2010) *Quarterly Reviews of Biophysics* **43**, 333-371
- [56] Valentina Tozzini and J. Andrew McCammon (2005) *Chemical Physics Letters* **413**, 123-128
- [57] Gergely Toth, Attila Borics (2006) *Journal of Molecular Graphics and Modelling* **24**, 465-474
- [58] Tozzini et al. (2007) *Journal of Structural Biology* **157**, 606-615
- [59] Rosemberg O. Soares, Paulo R. Batista, Mauricio G.S. Costa, Laurent E. Dardenne, Pedro G. Pascutti, Marcelo A. Soares (2010) *Journal of Molecular Graphics and Modelling* **29**, 137-147
- [60] M. A. S. Perez, P. A. Fernandes, and M. J. Ramos (2010) *The Journal of Physical Chemistry* **114**, 2525-2532
- [61] Seema Mittal, Yufeng Cai, Madhavi N. L. Nalam, Daniel N. A. Bolon, and Celia A. Schiffer (2012) *Journal of the American Chemical Society* **134**, 4163-4168
- [62] Gary S. Ayton and Gregory A. Voth (2010) *Biophysical Journal* **99**, 2757-2765
- [63] Tozzini, Rocchia, McCammon (2006) *Journal of Chemical Theory and Computation* **2**, 667-673
- [64] Tang, Ndassa and Summers (2002) *Nature Structural Biology* **9**, 537-543
- [65] Kuzmanic and Zagrovic (2010) *Biophysical Journal* **98**, 861-871
- [66] Federico Poloni *Diversi approcci al problema della classificazione*
- [67] Chen, Yao, Liu, Luo, Zhou, Jiang (2012) *J Mol. Model* **18**, 3153-3160
- [68] Trovato, Nifosì, Di Fenza, Tozzini (2013) *Macromolecules* **46**, 8311-8322
- [69] A.Bochicchio, V.Tozzini 'Multi-Scale Simulation of β_2 Microglobulin' in preparazione