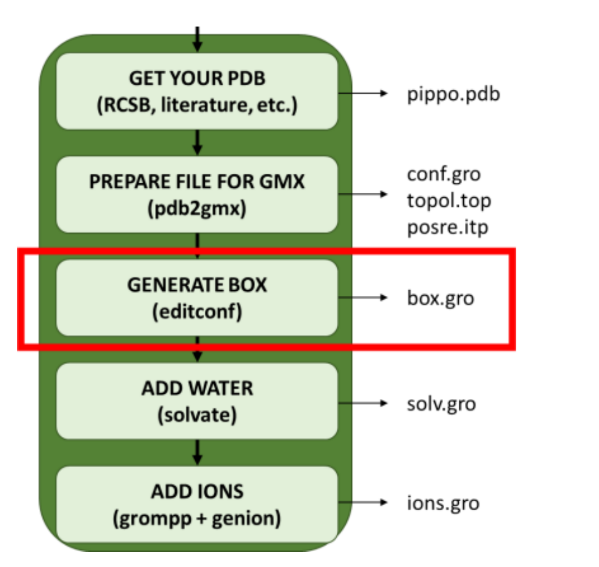
**TUTORIAL 1**

Partiamo pulendo la shell e nella cartella in cui siamo dobbiamo avere il file pdb contenente la proteina e i template che ci serviranno.

Quando è pdb le lunghezze dei legami sono in Armstrong, la riga CRYST1 è il box, cubico perché gli ultimi 3 sono gli angoli del box, i primi tre sono le grandezze del box nelle tre dimensioni in nm.

Nel pdb troviamo la descrizione della struttura:

* Numero atomo
* Nome dell’atomo
* Nome dell’amminoacido
* Numero dell’amminoacido
* Posizioni in x,y,z (armstrong) rispetto allo 0 del file, poi ci interesseranno le relative
* 1 ?
* 0 è correlato ai b factors, qui sono tutti zero perché derivano da una simulazione e non da una cosa sperimentale e ci danno l’idea dell’incertezza della posizione dell’atomo
* Tipo atomico

Un pdb termina con TER ENDMDL

*Per visualizzare un file pdb possiamo usare VMD, utilizziamo spesso cartoon che ci permette di vedere le strutture secondarie. In realtà VMD non sa come sono connessi gli atomi ma controllando le distanze tra gli atomi è in grado di calcolare automaticamente chi è legato a chi e di rappresentare la struttura.*

*Se proviamo a modificare le coordinate di un atomo e cambiamo rappresentazione, si vede come l’atomo non sia più legato alla struttura perché VMD lo vede lontano e non lo integra all’interno della proteina.*

NEL FILE PDB O GRO COMPAIONO SOLO LE COORDINATE MA NESSUNA INFO SULLA CONNESSIONE TRA LE PARTICELLE (sono nel top), in realtà nel pdb che vengono dagli esperimenti ci sono alcune info, per noi no.

**FORCE FIELDS (#pdb2gmx)**

Ci manca ora il modello, che identifica il campo di forze associato a quella struttura, che generiamo con pdb2gmx ma dobbiamo assicurarci che il file sia pulito ovvero non contenga molecole d’acqua.

*Potremmo farlo manualmente: prendiamo il file con WSH, facciamo:*

*cat nomedelfile.pdb | grep SOL | wc -l ci dice quante molecole d’acqua ci sono.*

*Se le vogliamo togliere facciamo cat nomedelfile-WSH.pdb | grep -v SOL > nomedelfile\_senzacqua.pdb*

**GENERAZIONE DI UN INDICE (#make\_ndx)**

Possiamo altrimenti usare un indice: gmx make\_ndx contiene un file che contiene dei gruppi in cui classifica gli atomi; l’indice è un file che contiene gli atomi clusterizzati per gruppi

Gmx make\_ndx -f nomedelfile-WSH.pdb -o start\_ndx.ndx

Gli sto dicendo di crearmi un indice che salva in start\_ndx a partire dal file\_con\_acqua.pdb

Quando noi lanciamo gromacs fa una prima suddivisione interna perché ha internamente un database e sa riconoscere diversi gruppi di atomi tra cui gli amminoacidi che mette nel gruppo proteina. Poi ci dice in che gruppo ha messo cosa: gruppo 0,1,..

Es:

* Protein-H è tutta la proteina togliendoci gli H associati, può essere utile
* C-alpha sono 16 e ci dicono quanti residui abbiamo
* …

Possiamo cancellare un gruppo con il comando ‘’del 9’’, attenzione perché i gruppi si rinumerano in seguito ad una delezione. Però ogni volta me li ristampa e posso cancellare. Posso anche eliminare un gruppo di gruppi ‘’del 5-10’’.

Posso anche creare dei nuovi gruppi, semplicemente inserendo la categoria (r=residui, a=atomi) e il numero: ad es. a 100-200 mi crea un nuovo gruppo in fondo a quelli che ho già in cui inserisce gli atomi dal 100 al 200

Possiamo anche unire i gruppi: se voglio tutta la proteina (gruppo 1) ma non il residuo 43 (che ho precedentemente salvato nel gruppo 10) posso dirgli ‘’1 & !10’’

& (and) vuol dire intersezione, prende entrambi gli input e prende solo quelli in comune

| (or) vuol dire unione

* Per salvare si fa q

**CONFIGURAZIONE (#editconf)** sto facendo la pulizia della proteina dall’acqua. In un modo alternativo a quello illustrato precedentemente

Una volta creato il gruppo dobbiamo pulire il pdb editando la configurazione usando gmx editconf; editare significa modificare le posizioni (traslare, ruotare ecc) ma anche aggiungere o togliere pezzi, costruire un box.. volevo togliere l’acqua, gli do in input il file con l’acqua e l’indice

Gmx editconf -f nomefile-WSH.pdb -n start\_ndx -o nomefile\_noacqua.pdb

Non gli sto dando indicazioni quindi lui mi chiede cosa voglio in output, io voglio il gruppo 1 che è la lista delle proteine senza acqua. Quindi ho finalmente il mio modello di proteina senz’acqua (lo avrei potuto ottenere cancellando a mano l’acqua come illustrato sopra)

Se vado a rappresentarlo con **VMD** mi serve un .gro mentre ho ancora un .pdb

Ora mi servono le connessioni ovvero i parametri del campo di forz (indica l’insieme dei parametri legati alla definizione dell’energia potenziale del sistema) e, nel nostro caso sono all’interno di un programma di gromacs che ha la capacità di creare all’interno di un database quali interconnessioni ci sono à pdb2gmx, mi restituisce un .gro

Convertirà le coordinate in un formato che piace a gromacs che è .gro e poi cercherà altre info, dando in output un file .top che è proprio quello che mi dice come è fatto (interconnessioni ecc)

In input prende solo il file pdb, eventualmente anche l’indice ma non ci serve; in output ci darà la topologia e il file di position restrains che ci servirà durante la fase di riscaldamento

Pdb2gmx -f nomefile\_noacqua.pdb -o start.gro -p topol.top -i posre.itp -ignh -heavyh

-ignh ignora gli idrogeni, se li va a ricalcolare tutti lui sulla base di quello che è il database. Così che la loro posizione venga ottimizzata, ci permette di avere un energia potenziale di partenza un po’ minore.

-heavyh serve a rendere più pesanti gli idrogeni per aumentare la dinamica perché sennò sarebbero troppo leggeri (1 Da), li appesantiamo così vibrano meno, diminuisce la vibrazione e diminuisce la frequenza di vibrazione del sistema da cui dipende il time step

Dando l’invio mi chiede che Force Field scegliere, anziché il classico AMBER (e poi TIP3P per l’acqua) prendo un GROMOS e poi il modello dell’acqua che scelgo SPC. Ho preso il GROMOS anziché l’AMBER, ci fa andare più veloce che risparmia un po’ di idrogeni perché li mette solo all’interno del tipo atomico. Amber va un po’ meglio perché è esplicito ma se ho un file molto grande meglio questo.

Se stampo il file .gro vedo che non ho più idrogeni vicino ai carboni alpha perché l’idrogeno è trattato come un tipo atomico, solo inglobati nel CB.

SE NOI CAMBIAMO IL CAMPO DI FORZE CAMBIA IL NUMERO DI ATOMI

A questo punto nella nostra cartella dobbiamo avere:

* Posre.itp, file del position restrains
* Start.gro, file gro che contiene le info della proteina
* Topol.top, topologia del sistema con info su configurazione, legami, struttura secondaria

Se stampiamo il gro (nano start.gro), contenente le coordinate, vediamo che ci sono delle differenze con il pdb

Se stampiamo la topologia .top contiene tutti quei parametri che raccontano qual è la natura della molecola che consideriamo e com’è fatta e organizzata (moleculetype), parametri che la definiscono. Ci dice che è una proteina, poi c’è la sezione *atoms* che ci dice la natura della struttura ma non come sono posizionati.

Ci fornisce info importanti quali la carica in .. e la massa in Dalton, che prima non c’erano

Poi c’è la sezione sui diedrali e dei legami e che funzione usa per legarli. Richiama poi i parametri quali gb2 che vengono definiti all’interno di un altro file.

C’è anche il campo di forza dell’acqua: due bond e un angolo, non ci sono diedrali perché sono solo 3 atomi

*Ricapitolando nella topologia ci sono tutti i parametri del campo di forza specializzati per il modello che gli ho dato, i termini vanno tutti a finire nella funzione di energia potenziale sommatoria di termini di bond e non bond.*

*Nella TOPOLOGIA non compare nulla che faccia parte dei termini di non legame che compariranno solo durante la simulazione, mentre quelli di legame li so a priori della simulazione. Gli unici termini relativi a termini di non bond presenti sono i parametri di VdW, che sono i fattori sigma ed epsilon (c6 e c12, parametro di collisione e di energia).*

**BOX (#editconf)**

Abbiamo finalmente il nostro modello che dobbiamo mettere all’interno di un box per poter definire il dominio e i suoi estremi.

gmx editconf -f start.gro -o box.gro -bt cubic -d 0.8

Sto facendo un box cubico, l’importante è che la minima distanza dalla fine della proteina alla fine di un lato del box sia almeno 0.8.

\*editconf comando già usato per togliere acqua

Mi fornisce anche le coordinate del centro del box, ci da il volume e ci ha spostato la proteina perché gli viene meglio. Infatti se gli carico dentro le due proteine (quella nuova del box e quella senza box) vedo che sono spostate

Box.gro è il nuovo file contenente l’info della proteina inserita nel box e senza acqua

**NUOVO INDICE**

Dopo, devo usare un nuovo indice perché ho bisogno di un indice che non contenga gli idrogeni che ho eliminato, quindi i numeri non corrispondono più, il vecchio contiene i numeri degli atomi di ogni gruppo contenente gli idrogeni ma ho usato un modello GROMOS che mi ha levato degli idrogeni integrandoli nel tipo atomico quindi l’index vecchio è tutto sfalsato e devo buttarlo e farne un altro nuovo.

**ACQUA**

Gmx solvate -cp box.gro -cs spc216.gro -o sol.gro -p topol.top

Gli diamo anche la topologia perché ci sono info sull’acqua

Il software aggiunge l’acqua al sistema copiando all’interno del box il file spc216 che viene copiato all’interno del volume del box e poi cancella le parti che si sovrappongono alla proteina.

Se usiamo l’opzione -shell (ci conviene risalvare in un nuovo file output -shell.gro), si aggiungono meno molecole d’acqua perché sol.gro ha riempito il box con l’acqua dove ha potuto, ma se usiamo shell invece anziché riempire tutto va solo a mettere un sottile strato d’acqua intorno alla proteina.

~~-radius (creiamo nuovo file per vedere differenza) mettiamo -radius 0.07 parametro che usa per identificare se c’è sovrapposizione ovvero se deve togliere dell’acqua. Dovrebbe permetterci di mettere più o meno acqua ma non funziona~~

La funziona -vel o -novel ci permette di considerare o meno le velocità dell’input del soluto o del solvente nel caso avessimo già una dinamica.

Dobbiamo ora chiederci se il sistema ha carica neutra, dobbiamo simulare per capirlo à ci serve un file tpr

Abbiamo quindi generato il nostro sistema:

* File gro
* File top, contiene sia la top del soluto che del solvente. In topol.top, infatti, c’è in fondo anche il sol e non solo la proteina.
* Indice da aggiornare
* …

**MINIMIZZAZIONE molecolare** , (#grompp) vogliamo avvicinarci al minimo di energia

Necessitiamo di una file di parametri che ci permette di fare la minimizzazione di energia. Abbiamo un file template.mdp che ci permette di modificare tutti i parametri. Una prima parte da non modificare.

Comm-mode serve per quando si lavora in vuoto anziché in acqua

Compaiono anche indicazioni sulle interazioni ovvero sul cutoff, ma noi non cambiamo nulla (solo teoria). Pbc considera il trattamento delle condizioni al contorno, nstlist è la frequenza di aggiornamento della lista, rlist è il cutoff e deve essere omogenea rispetto al cutoff di Coulomb o VdW. Poi abbiamo il trattamento di interazioni a lungo raggio (Coulomb, ovvero che guarda il cutoff che è il metodo peggiore). Idem Vdw.

Copiamo il file template (cp template.mdp 1em.mdp) cosi possiamo modificarlo per la simulazione, lasciando la prima parte invariata

integrator steep (stiamo dicendogli di fare una minimizzazione con il metodo steep descent)

Togliamo dt

Nsteps mettiamo -1 che significa va all’infinito finchè non converge per altri motivi (emtol: forza massima, o precisione della macchina non riesce a distinguere la diminuzione di energia perché troppo piccola e si ferma)

**Emtol** (forza massima che accettiamo sull’atomo) in kJ/mol/nm che mi fa fermare la simulazione mettiamo 10

**Emstep** (criterio di movimento) mettiamo 0.01 nm

OUTPUT CONTROL OPTIONS

5 steps per salvare la traiettoria

5 steps per salvare le velocità

5 steps per salvare il log, ovvero mi da info su errori, instabilità del codice ecc

5 steps ogni quanto salva le energie del sistema, ci serve per la minimizzazione e la dinamica

5 numero di energie da calcolare (?)

Tra che gruppi calcolare le energie energygrps = Water non-Water; ad esempio calcolo le energie di VdW tra i gruppo con e senz’acqua. Nel caso di simu in vuoto non mettiamo nulla

Cancelliamo tutto tranne temperature coupling, steep non lo legge nemmeno quindi possiamo non cancellare

L’annealing c’entra con la temperatura e la cancelliamo

Generate velocities: no;

In gromacs per lanciare la simulazione si usano due comandi:

* Il preprocessatore: gmx grompp -f 1em.mdp -c sol.gro -p topol.top -o 1em.tpr -maxwarn 2
* ?

Il tpr, che è l’output, contiene la simulazione

Quando lanciamo ci da il warning del fatto che la carica sia 7 e non neutra. Dobbiamo risolvere questo problema aggiungendo ioni. In più ci da un altro warning in cui ci spiega che abbiamo usato un modello non corretto GROMOS, legato al cutoff ecc.

Per eliminare gli warnings e fallo andare avanti aggiungiamo -maxwarn 2 che ci permette di vedere comunque gli output.

Ci crea in automatico anche mdout.mdp: è il riassuntone e ci richiama tutti i parametri che abbiamo usato, con anche tutto ciò che non abbiamo messo.

**NEUTRALIZZAZIONE DELLA CARICA (#genion)**

Dobbiamo però sistemare il problema della carica, inserendo degli ioni neutralizzanti con il comando

gmx genion -s 1em.tpr -neutral -conc 0.15 -o ions.gro -p topol.top

-neutral e -conc 0.15 ci permettono di inserire ioni fino ad avere una concentrazione di 0.15 mol/l ma in modo tale anche da avere carica neutra

!! ogni volta che modifico la composizione del gro, in questo caso sto aggiungendo ioni, devo anche aggiornare il top perché sennò non c’è più corrispondenza

Ci chiede in che gruppo andare ad inserire gli ioni e gli diciamo SOL

Vediamo come si è aggiornato il gro, guardando direttamento ions.gro che è quello con anche gli ioni che sono inseriti in fondo al gro originario

La topologia ci mostra quanti ioni sono stati aggiunti

Faccio q e salvo

Per essere sicuri che gli ioni siano stati inseriti facciamo un nuovo indice con gmx makendx -f ions.gro

!! water non contiene gli ioni, dobbiamo o creare un gruppo cumulativo o chiamare entrambi i gruppi

In teoria dovremmo accoppiare gli ioni con l’acqua ma di fatto nel progetto si fa water non water e bom (chiarisci)

**SIMULAZIONE 2 di minimizzazione dopo aver neutralizzato**

Riproviamo a fare la simulazione con gmx grompp ma la topologia è stata aggiornata e quindi ci sono anche gli ioni che ci permettono di avere carica nulla e anche il file di input non è più sol.gro (box+proteina+acqua) ma usiamo ions.gro che contiene anche gli ioni. Verifichiamo che la carica sia nulla dagli warnings

Gmx grompp -f 1em.mdp -c ions.gro -p topol.top -o 1em.tpr

Non ci sono più note sulla carica, quindi nulla

**SIMULAZIONE DI MINIMIZZAZIONE**

Vogliamo lanciare la **simulazione di minimizzazione energetica** che utilizza lo steepest descent (info contenuta nel mdp) : gmx mdrun -s 1em.tpr -v -deffnm 1em -nt 2

-v output verboso, ci dice quello che fa a parole mentre simula

-deffnm se non gli do niente me lo chiama come vuole, io lo chiamo 1em

Se non siamo in acqua ci mette molto meno steps a minimizzare perché ci sono meno molecole da riarrangiare ma l’energia viene più alta; ma quella in acqua ha un valore molto più negativa perché ci sono molte più interazioni da calcolare in quanto quelle di coulomb sono tutte attrattive quindi negative.

Mi restituisce un sacco di file che mi dicono come avvenuta la simulazione

.trr l’insieme di tutte le posizioni ad ogni step delle particella della minimizzazione, ovvero la traiettoria. Quando lavoreremo in dinamica mi dirà le posizioni delle particelle ad ogni istante di tempo

.log come lavora il software, errori ecc

.edr contiene info su energie pot,cinetiche, temperature ecc

**CONFORMAZIONE**

Vmd 1em.gro 1em.trr

Per vederlo, comando vmd, devo caricare prima il 1em.gro poi il 1em.trr, se lo faccio in verso opposto non è in grado di leggerlo perché è solo una traiettoria non ha nessuna info sulla molecola.

Si vedono i micromovimenti che la struttura fa per raggiungere il minimo locale mediante un riaarrangiamento della struttura.

Per vedere un’energia specifica usiamo gmx energy, uso in input .edr

Gmx energy -f .edr -o pot.xvg

Con xmgrace possiamo visualizzare l’energia potenziale

**POTENZIALE**

Gmx rms -f 1em.trr -s 1em.tpr

Poi gli diciamo 3 (c-alpha, mi toglie la rototraslazione rigida controllando che i c-alpha siano invariati) e 4 (backbone, mi calcola l’rmsd solo del backbone ovvero la catena primaria senza gruppi laterali: ovvero di quanto sono cambiati le posizioni da quella iniziale).

Non legge il gro ma vuole il trr perché fa una media pesata sulla massa e nel gro quest info non c’è e non la sa calcolare, mentre nel trr c’è anche il top che ha le masse

Mi chiede per chi fare il RMSD, secondo input, (spostamento di ogni particella dalla posizione iniziale, sommatoria mediata), se però la proteina rototraslasse e basta il rmsd sarebbe nullo. Mi toglie la rototraslazione rigida e gli faccio calcolare questo spostamento per la proteina (??)

Per visualizzare il RMSD, il gmx rms ci ha creato un file .xvg che andiamo ad aprire con xmgrace rmsd.xvg

Il massimo valore di spostamento è 0.04 nm vuol dire che non sta esplorando nulla perché siamo sotto l’Amstrong, quindi durante la minimizzazione non c’è alcuna esplorazione ma solo un riarrangiamento che ci permette di raggiungere il minimo.

\*l’operazione di minimizzazione dell’energia potenziale è stata fatta in due step (grompp e poi mdrun), prima il programma processa gli input:

* coordinate nel gro (parte dinamica che si modifica durante la simu)
* la topologia (coordinate statiche che non si modificano, parametri della V)
* mdp che ha i parametri per la simulazione