Bericht aan anton:

* Het is wel een beetje oneerlijk dat er steeds meer informatie wordt gegeven naar het einde vn de deadline.
* Ik begrijp wat jullie de vorige keer zeiden: hoe meer struggle, hoe meer je leert. Dat ervaar ik zelf ook zo en dat is goed.
* Maar bijv: je bent best veel tijd en aandacht kwijt geweest aan het vak en je bent best aan het struggelen geweest. Top. Je vind redelijke antwoorden op de vragen en bent klaar. Een aantal dagen voor de deadline, want je moet ook werken (geld) en afstuderen (ook nodig). Wellicht krijg je voor dit wat je hebt ingeleverd een 6. Nou komt er op de dag voor de deadline een vragen uur en slides online, waar veel meer informatie wordt gegeven. Je kunt de dingen die je hebt geleerd nog beter aanvullen en zou je opdracht omhoog kunnen tillen naar een 7 of 7,5! Echter heb je hier geen tijd voor. Je hebt precies die dagen voor de deadline ander verplichtingen. Hierdoor blijf je met je 6 zitten. Nu vind ik het persoonlijk niet z’n probleem of groot verschil. Een 6 of een 7, allebei prima. Maar het maakt wel net uit wanneer je een 5 hebt voor je toets en je met een assignment net je gemiddeld eomhoog kan tillen.
* Ik begrijp dat jullie niet meteen alles voor willen zeggen. Maar hoe veel struggelen is eerlijk en wanneer is het delen eerlijk? Willen jullie hier alsjeblieft een keer over nadenken?
* Verder vond ik het vak best intens. Ik heb 2 keer gehuild ☺

Structural bioinformatics

Assignment 3

Lianne van Ruijven

# Notes

chromacs —> anton made this

We use umbrella sampling to create a free energy landscape

have to use the VU server. Do it individual

---------------------------------------

Goal of the practical assignment:

Find the free energy when a band between two atoms is broken..

1VET protein.. has interaction between two parts.. sampling difference between the states..

calcute free energy of different coformations..

we want to investigate the difference between the states where the A and B does not interact and do interacts.

We want to investigate the interaction between A and B..

## **Popular Answers (1)**

[](https://www.researchgate.net/profile/Dmitri-Kireev)

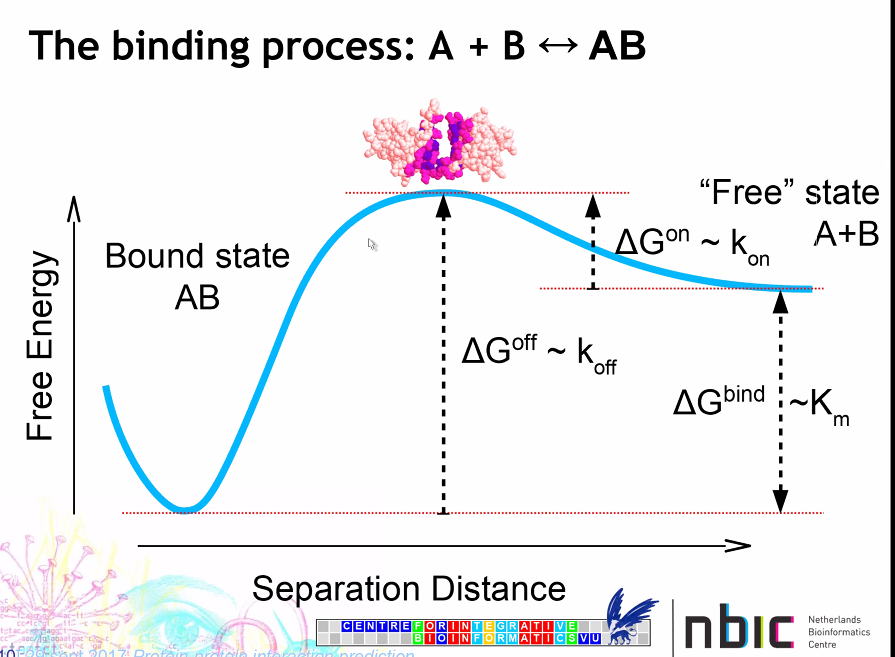
15th Jan, 2018

[Dmitri Kireev](https://www.researchgate.net/profile/Dmitri-Kireev)

University of North Carolina at Chapel Hill

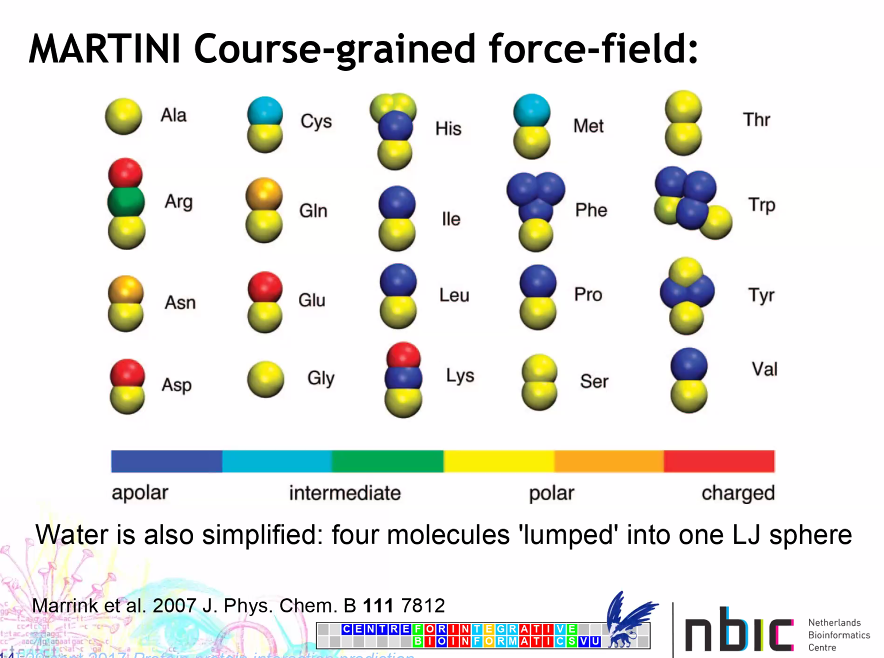
I'll make an attempt to provide a yet simpler view. The purpose of any molecular dynamics (MD) simulation is to sample (all) possible states in which a molecule of interest may exist. Based on this sampling, probability (free energy) for the molecule to be in any of these states can be readily calculated. Very (too) often, certain states of the protein are separated from others by extremely high energy barriers. Sometimes, it would take years, if not centuries, of conventional MD simulations to walk through all molecular states. Umbrella sampling (among other techniques) allows to accelerate the sampling by "flattening" those hills and ridges, which prevent MD from accessing certain states. In umbrella sampling, the energy landscape is "flattened" through adding artificial "umbrella" potentials that are supposed to "mirror", and thus annihilate, the real barriers. However, it would be difficult, if not impossible, to make an umbrella potential account for all degrees of freedom in the system (there are a few thousand of them for a typical protein, even if the solvent is neglected). Hence, the umbrella potential involves only a few (most often, one to three) degrees of freedom, often called collective variables or reaction coordinates. Sampling of a system is considered complete when it has "visited" all values of collective variables more than once (that is, a number of times required for an accurate and unbiased calculation of state probabilities).

Lecture Anton 9 maart:



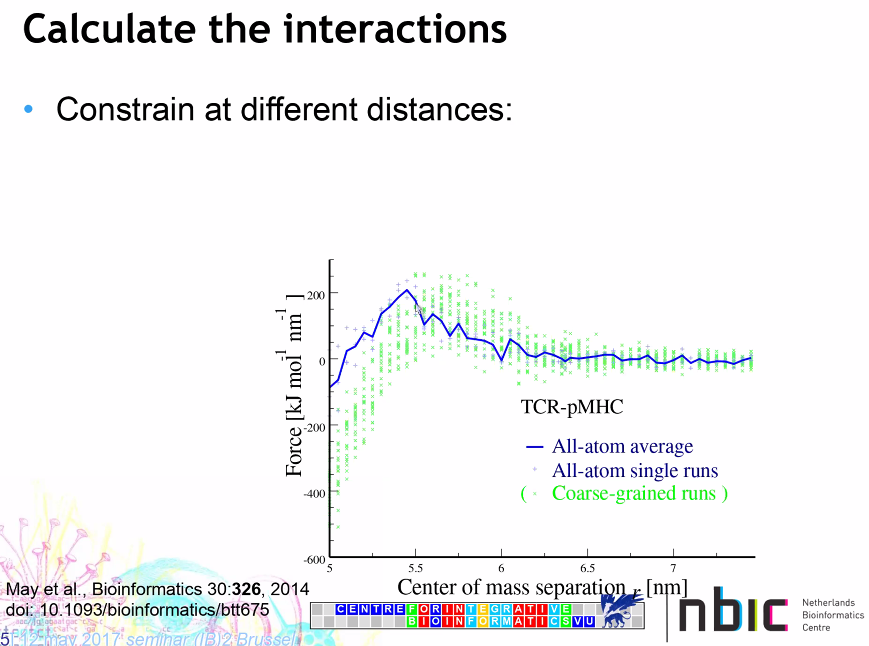
But being bounded is not always the most variable state, for example in a high temperature.

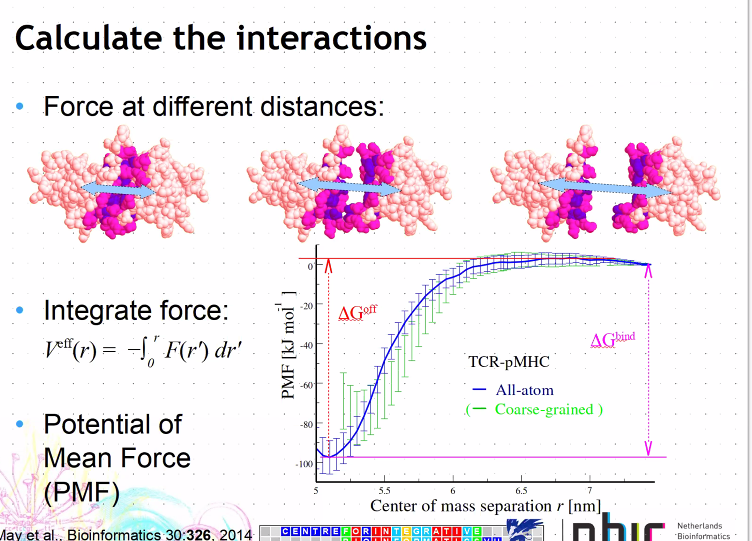
We used course grained force fields, Makes simulations 10 to 20 times faster:



How do you calculated the free energy landscape?:

* First you keep the two proteins at a certain distance and you calculate what Force is needed to keep the proteins at this distance.
* So the graph below is not like the regular free energy landscape graphs. This is only the force needed to keep the two proteins at a certain distance.
* PMF is now the average force needed per distance..
* PMF can be seen as the free energy





# Deliverables

Question 1

The technique used in the May et al. (2014) paper for calculating the PMF is constraint force integration.

→ Describe the essential difference between constraint force integration and umbrella sampling.

* The F mean (which is a function of the reaction coordinate), is calculated between a window of the reaction coordinate, when using umbrella sampling. For constraint force integration, the Fmean is calculated for specific values of the reaction coordinate.

→ Name two main assumptions that need to hold for these methods to work correctly (only one is explicitly mentioned in the paper).

* Number of water molecules assumed at an certain volume (or having a constant N (number of particles), V (volume), E (energy))
* Assumes equilibrium

# Starting up

The 1VET.pdb file contains coordinates.. how do you get the kinetic energy from a pdb file?

## Getting the necessary files

* For the 1VET structure: there is a chain A and chain B!
* Je kan zien welke delen elkaar raken.. is volgens mij de betastrands die dicht bij elkaar liggen

## Generating the topology

In this part:

* the pdb file is checked to be complete.. so are all atoms present in all amino acids.. The file 1vet\_add.pdb has alle the atoms that would be missing, added.
* Then the potential energie from the coordinates is determined.

done

## Setting up the system

Set up the size of the system.. so putting the system in a box of 1.2 nm around the molecule and it is centered in the box.

Conf.gro contains our molecule centered in a box

done

## Energy minimizing the system

We want to minimize the energy because… try to stabilize the structure. 🡪 The goal of energy Minimization is to find a set of coordinates representing the minimum energy conformation for the given structure.

* Output file Em/fit.pdb contains both structures, means: 1VET before and after minimizing the energy. Difference is super small, you cannot really see any difference. The super long loop in the protein connects the two chains.. this happens because the difference between the two chains is not that clear.. so this is an artifact. Chromacs doesn’t know about different chains..
* 0.098 is wat Fabienne krijgt. Dit is best wel laag.. dus komt wel overeen met de 3D model in chimerea.. dat de rmsd > 0 komt waarschijnlijk door de lange loop in het midden.

done

# Umbrella sampling

## Coarse-graining

Oke done, is just explenation

## Distance

Distance from c1 to c2 is 23.535

For the umbrella sampling:

* minimal is 23.535 armstrong = ongeveer 2.3 nm
* max ?? paper 4.44 as max

## Equilibrations

Ik heb range 2.2:4.5 gekozen

python3 ../src/Pull.py -f 1VET-em.pdb -c A:B -d 2.2:4.5:0.1

Lager dan 1.7 gaat niet zo lekker, omdat dan de twee proteïnes te dicht bij elkaar komen.

Question 2

We are sampling with the umbrella sampling. So each sample is a starting configuration in a box filled with water for a MD simulation. Why do we run short simulations to sample?

* so the water can adjust to interacting with the proteins and so we know the starting configuration is in equilibrium. This is also a good check to see if running the MD simulation is possible.

The difference between EM and MD is:

* The goal of energy Minimization is to find a set of coordinates representing the minimum energy conformation for the given structure. This energy conformation is the potential energy. So in EM, only the bonded (bonds, angles, dihedrals and impropers) and non-bonded (coulomb, van der waals) forces can be adjusted. This won’t have large effects on the secondary structure of a protein and the changes won’t be large.
* MD takes potential (coordinates) and kinetic (velocity) energy into account. An MD simulation without constraints can totally unfold and fold a protein, so has a large effect..

## Production simulations

Why did some distances fail?

2.2 and 2 and 3.3 failed for Fabienne.. you can see this at that there is no folder for this distances.

Anton doesn’t really know.. it also happens in the paper.. maybe this happens because the proteins are in a box.. maybe at these distances the interaction is on the other side of the box… 2.2

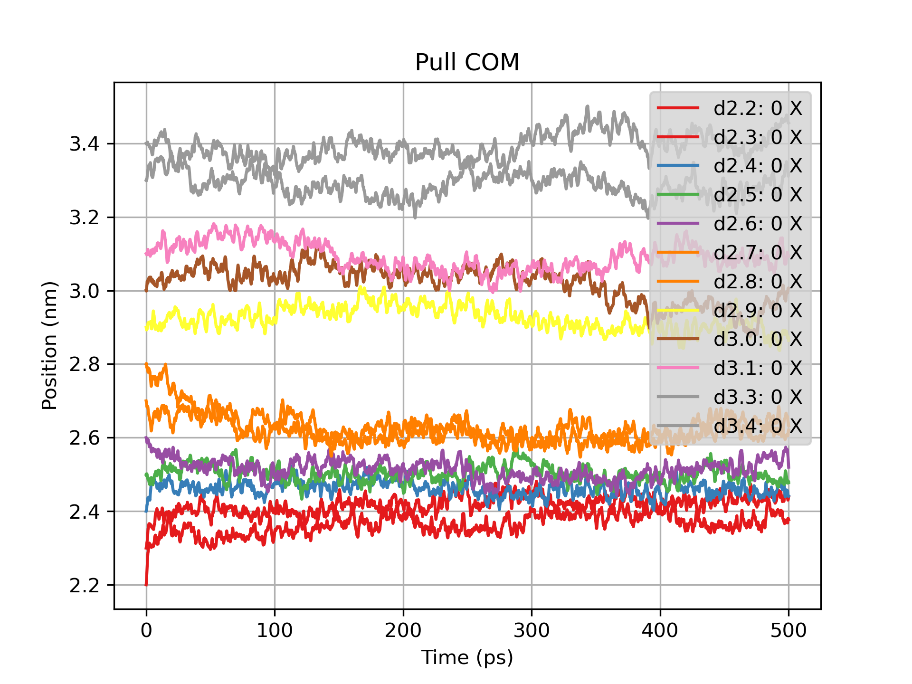
DONE!

# Analyzing the results

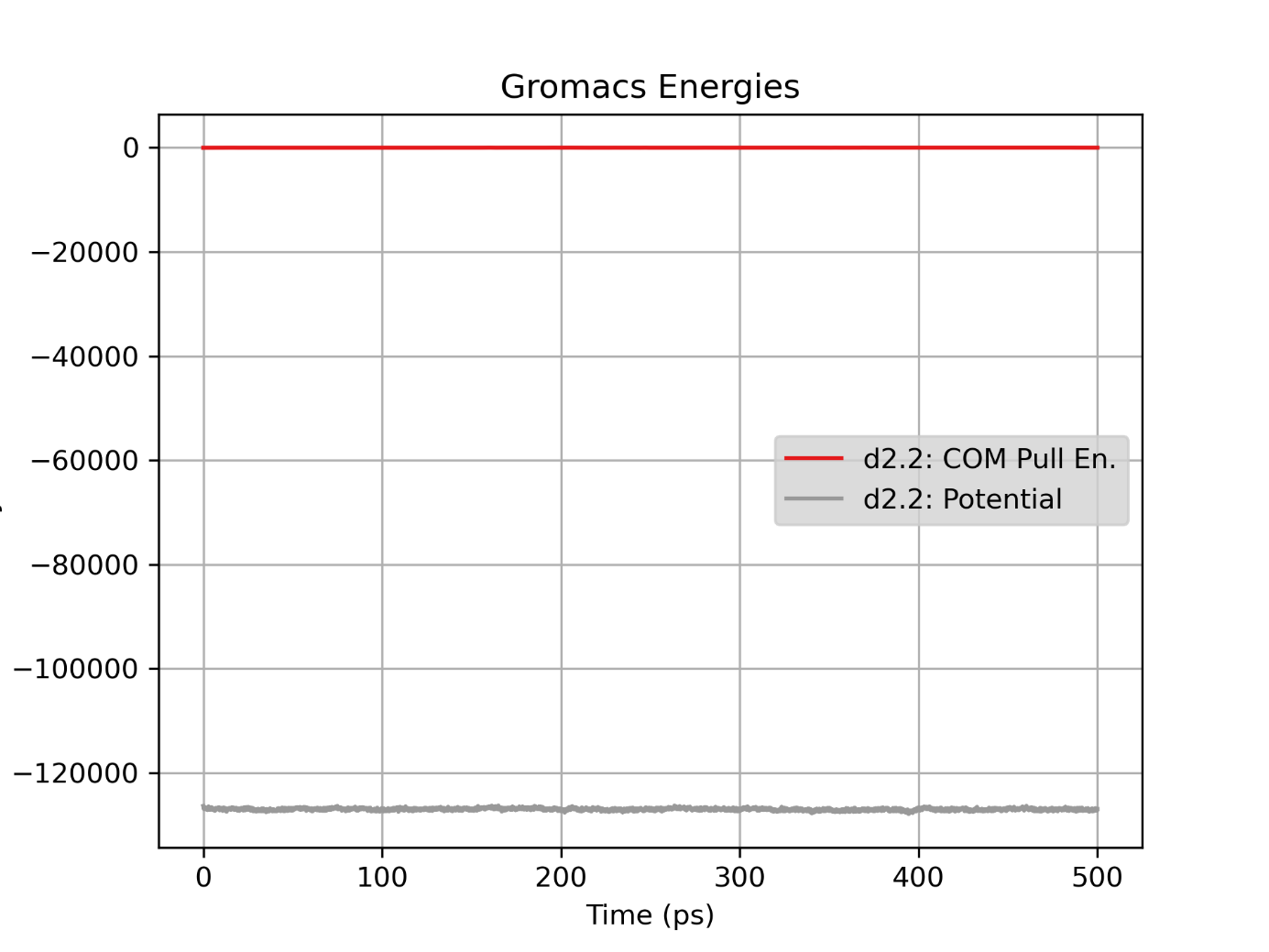
## Analysis according to umbrella sampling

### Extracting energies and distances

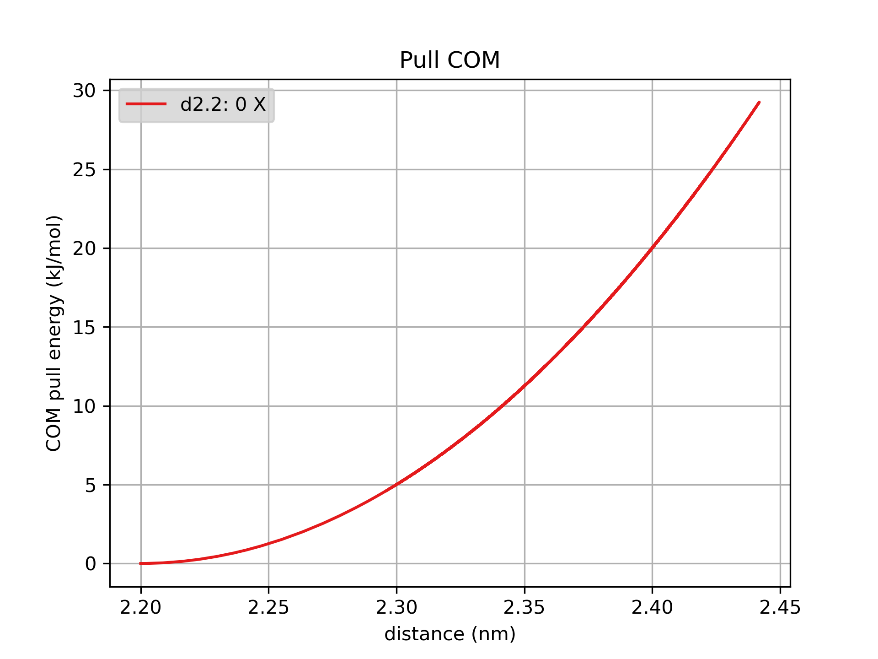
Center of mass distances as function of time:



Energy potential of sample with distance 2.2. COM pull En = umbrella potential energy and Potential = total potential energy.



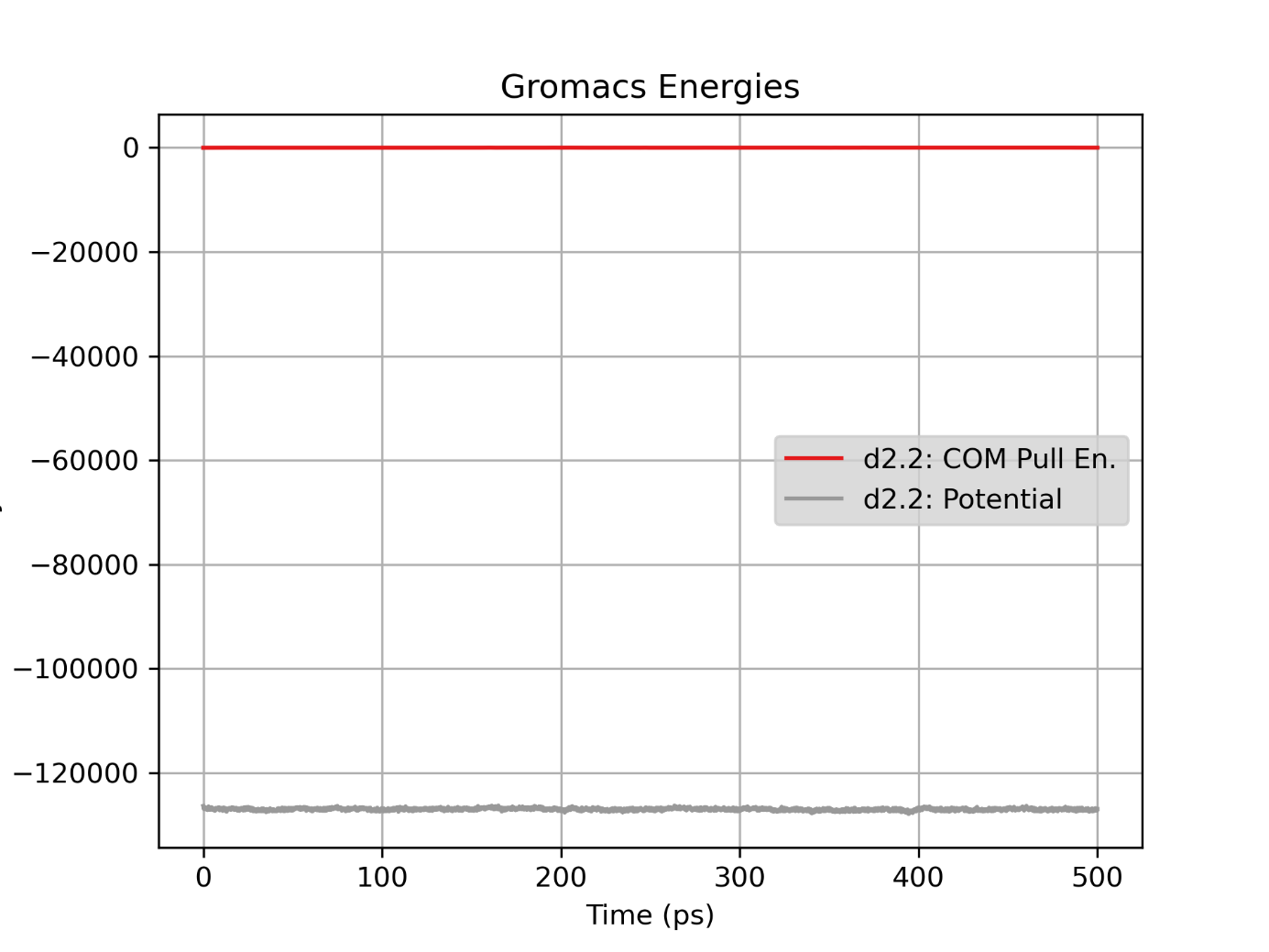
Question 3



What figure would you expect for an infinitely long sampling?

Question 4

Dit gaat niet goed, om antwoord te geven op de vraag:



Question 5

### Unbiasing the umbrella potential

## Automated analysis using WHAM

Question 6

Question 7

Question 8

Question 9

Question 10