

Name:

Liam O'Sullivan

Title:

Modelling molecular recognition and aggregation in nano-DNA mesophases

Supervisors:

Matt Glaser, Yves Lansac, Zach Smith

Introduction

In November 2007 people from the liquid physics group at the (1) University of Colorado and University of Milan showed that, short B form DNA (sDNA) of length as short as 6 base pairs form liquid crystals when put in saline water. This discovery has implications for the formation of life and it is reasonable to assume salt was present in this process. (I will explain later on how it has implications for the formation of life.)

DNA is chemically made up of a backbone made of sugars and phosphate groups joined by ester bond. Attached to each sugar is one of four types of molecules called bases. It is the sequence of these four bases along the backbone that encodes information.

It has been known for around 60 years that DNA forms a liquid crystal phase when put in water. However, Bolhuis and Frenkel (2) made a complete computer simulated phase diagram for B-DNA (modeled as a repulsive rigid or semi flexible rod-shaped solute) which predicted that liquid crystals would only form if the number of base pairs exceeded 28, no matter how high the concentration. Therefore it was a great surprise to find that sDNA stacked end-to-end, forming rod shaped aggregates. (Fig. 1 shows an example of a “palindromic oligomer”.)

Fig. 1

DNA helix. Note that the strands are the exact same, but arranged so that one is upside down. This can be referred to as a palindromic oligomer.

These DNA strands stack upon one another and pack next to each other. The reason for this stacking is unknown. A popular belief is that it is due to the hydrophobic effect. The theory is since the DNA ends are hydrophobic, they want to minimize the surface area exposed to water and if they stack upon one another they can “squeeze” the water out. The DNA helices then condensate into liquid crystal droplets within the water, which physically separates them from the non-complementary DNA strands (the DNA strands which have not found a matching strand). Some pictures of the liquid crystals formed are shown in Fig. 2.

It is entropically favorable for the strands to align side by side. The reason for this is that the orientational entropy it has gained from aligning out weighs the positional entropy it has lost from aligning.

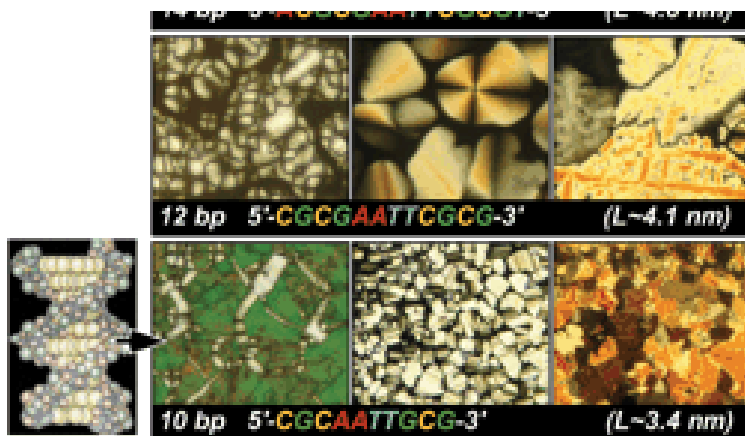


Fig. 2

The optical textures of the LC phase of a 12-base pair and 10-base pair sDNA solution in water obtained by depolarized light microscopy.

I am interested in modeling the behavior of the stacking of the DNA helices. I want to show, through computer simulation, that if two short DNA strands are aligned vertically with each other in water that it is energetically and entropically favorable for them to come together to form longer chains of DNA. I want to calculate the free energy from this interaction using a technique called umbrella sampling. I would also like to create a simple hard rod model of the DNA once I find the Free-energy.

It would be simple enough to calculate that it is energetically favorable for the DNA to stack, but showing that it is preferable entropically is harder because knowing what entropic interactions occur is a more complex issue. Therefore I use a molecular dynamics simulation program called Amber which has been previously developed. In doing this simulation I will be using some well known computational techniques and some less well known techniques, and I will elaborate on these methods accordingly. Using these techniques I will show that it is favorable for the DNA helices to stack on top of one another and align side by side with each other.

Method

Two DNA molecules were made and were then separated by a certain distance which we believed would be the optimum separation distance. The number of base pairs used is 12 base pairs for each helix and each helix is identical to start off with (one is just a duplicate of the other). Constraints of sodium and water were then added to the system. The salt is added to the system in order to neutralize the places of high negativity in the DNA molecule. As little water as possible is added to allow the runs to go faster. A rectangular box uses the least number of water molecules to surround a chain which is thin and long, where as an octahedron works better for smaller wider molecules. For the simulation in which the molecules were separated by a distance in the z-direction we used the rectangular box, and for the separation in the x-direction we used the octahedron.

A minimization procedure was then run for two reasons. Firstly an unequilibrated system is an unrealistic situation. Secondly the water does not feel the influence of the solute (DNA) or charges. Moreover there may be gaps between the solvent (Water) and solute and also possibly the solvent and the boxes edges. Such holes can lead to 'vacuum' bubbles forming, thus a careful minimization is run before slowly heating the system to 300K. The interactions between the particles are calculated according to Newton's equations of motion and using the verlet algorithm to calculate the speed of the particles. For the equilibration the Particle Mesh Ewald (PME) method is used. To prevent the outer solvent molecules from boiling off into space, periodic boundary conditions are used. This technique prevents the system from losing its particles and minimizes simulation time by copying the structure of a certain central box of particles and surrounding this box by the other copied boxes. Every time a particle leaves the central box the particle enters at the exact opposite position in the box (at the other side of the box) with the same velocity, thus ensuring no particles are lost.

There are two steps to the minimization procedure. First the DNA is held fixed using positional constraints, which keeps the DNA atoms in the same place. The positions of the water and ions are then allowed to minimize. In the second stage the whole system is minimized.

It is seen in the output of the minimization of the water and ions only, that the electrostatic energy drops very dramatically at the start because the water is re-orienting itself into its preferential bonds (water-water or water-DNA bonds.) The temperature of the system is raised to make it come to around room temperature (300K) to make it a realistic situation. Initially we keep the DNA molecules fixed for around 20ps, and then we heat the whole system up for 100ps. The kinetic and potential energies make an increase during the initial heating of the system then they both level out indicating the system is equilibrated. We can check that the structure is what we expect it to look like by using a program called VMD. Once it has been checked that the structure is acceptable, The simulation can begin.

The simulation we run uses a less well known technique called umbrella sampling. The Umbrella sampling technique requires a force be added between the two DNA strands. A spring is attached to the centers-of-mass of the two DNA strands and a suitable spring constant is chosen. The simulation is then allowed to run. The motion of the DNA strands obeys the equations of motion, but with the added effect that the water has on their motion. A video of the simulation can be run during the process of the simulation using a program called VMD. This allows us to check if the strands have done anything unexpected during the simulation. The data is recorded and histogram plots of the separation distance between the two strands are made. A time correlation function of the simulation is obtained and this is used to find the error in our simulation.

Discussion

In doing this simulation we hoped to find the Gibbs Free Energy of two strands of DNA interacting in water. We also hoped to create a hard rod simulation of the interaction between numerous DNA strands without the added effect of water. The first part of this experiment was achieved, however there was a very high error associated with parts of the curve indicating this simulation was not run for long enough. It was not possible to start the next part of our experiment due to time restrictions. Also not knowing the curve of the free energy would have stopped us from making a model which was accurate.

The free energy was obtained using a technique called umbrella sampling. Umbrella sampling is a form of Monte-Carlo simulation, except it allows you to sample regions which the Monte-Carlo method wouldn't sample enough normally. Umbrella Sampling allows you to sample only the areas of the curve you are interested in. How it works is the algorithm adds an extra probability to the region you are interested in.

In our model we had two duplex DNA strands, so we had to add a potential between them in order to increase the probability of the system staying in the region we wanted to sample. The potential was increased by adding a spring between the centers of mass of the DNA strands. It was possible to adjust the size of the region we measured by adjusting the spring constant of the spring. A higher spring constant gives us a narrower window, which means better accuracy, but there is a compromise and that is that there is then less overlap between the windows, and so more samples are needed.

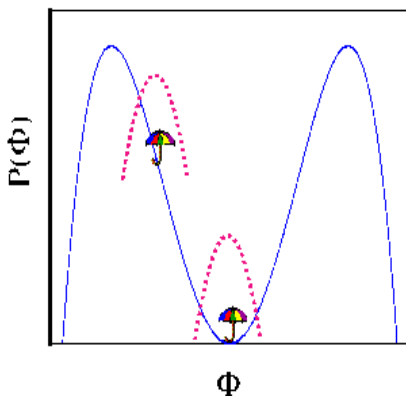


Fig. 3

In order to plot the graph of the Free-Energy two main programs were used. One is a histogram reweighting program used to measure how much the system fought the spring force in order to go back to its natural state. The other was a time correlation program we used to find how independent two events which occurred during the simulation were. The histogram reweighting program along with the time correlation program enables us to plot the graph of the Free Energy. Fig. 3 above shows how probabilities are added to certain regions.

The histogram plots below in Fig. 4 are the histograms of the various separations in the z-direction. We believe that more windows and possibly a stronger spring constant will be needed if we are to have an accurate curve for the free-energy because as can be seen from Fig. 6 the error bars associated with the

free energy are quite large as we go into regions where we did not take as many samples. In order to run longer simulations more processors will be needed, as the simulations run on our computers take too long. Right now on 4 processors in parallel it takes 70 hours to complete a simulation of 2 nano-seconds. It will only take 8 hours on 48 processors. It can also be seen from the correlation function in Fig. 5 that this simulation is not long enough as we see that the function still has not gone to zero. Even after going halfway through our simulation there are still no two independent events. Since most events in our simulation are correlated, it gives us a much higher error in the free energy obtained.

Fig. 5

Although we have found out that we need to run longer simulations, we have obtained very promising results from the data we have obtained. The region in which we sampled gave us what we expect to be the minimum of the free energy curve. We imagine that the rest of the curve will look something like Fig. 7. If we can find the free energy distribution for this interaction it would be possible to make a model of hard rods in order to see what happens on a macroscopic scale.

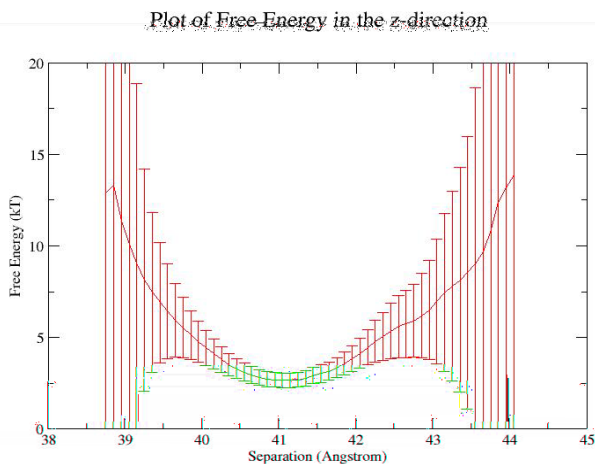


Fig. 6



Fig. 7

References

1. Michi Nakata et al, Science 318 (2007)
2. Peter Bolhuis and Daan Frenkel J. Chem. Phys 106 666 (1997)