Jodie Scrimshaw S3823172 December 4, 2023

Parameterising Martini 3 Coarse-Grained Force Field for B-DNA

Supervisor: Alex de Vries

Abstract

This study focuses on developing the Martini 3 force field for coarse-grained B-DNA, specifically targeting adenine and thymine residues. Utilizing GROMACS 2021.5, the parameterization process, based on all-atomistic (AA) mapped trajectories from B-DNA simulations, revealed considerable overlap in bond, angle, and dihedral distributions. Coarse-grained (CG) bonds were not able to replicate the narrow AA distribution as changing force constants caused system instability. Simulations of single strands (A or T residues) with the optimized force field displayed a loss of compact helical structure and disrupted base stacking. Clustering analysis comparing base-base interactions in solution showed slightly stronger interactions in the AA model (-7.5 kJ mol^{-1}) than the CG model (-5 kJ mol^{-1}). While the clustering analysis did not elucidate the absence of base interactions, the study severely underscores the importance of entropic driving forces in base stacking and base pairing. In simulations of CG single-strand DNA, severe elongation of the backbone was observed. When double-stranded CG DNA was simulated, the same occurred except significant base stacking was observed between alternating bases of each strand. This shows that the backbone strain is too high for base pairing to be feasible. According to a study, this issue is entropic in nature, explaining that the entropic penalty of backbone conformation must be compensated by the entropic gain of base pairing to minimising solvent interaction. Increasing the enthalpic driving force by making the bases have more favourable interactions did not compensate for this. In this study's concluding phase, the entropic driving forces within CG single strand DNA were maximised by modifying bead types in accordance with literature insights. Specifically, central base beads (SC6 and SC1) were rendered maximally hydrophobic (type C1), side beads (SC5 and SC2) were moderately less hydrophobic (type C4), and base-pairing beads (SC3 and SC4) were rendered maximally polar (type P6). Hydrophobic beads were additionally designated with 'h' labels to enhance attractive forces between bases. Following these adjustments, the single strand DNA exhibited notable base stacking and helical conformation, albeit with relative instability compared to all-atomistic simulations (AA). The DNA also displayed a proclivity for base pairing, attributed to a significant polar/non-polar discrepancy within the base, hypothesized to impede intercalation. To enhance model stability, future research should explore the delicate balance between backbone strain and base stacking interactions, emphasizing entropic driving forces.

https://fse.studenttheses.ub.rug.nl/id/eprint/31731.

Contents

1. Introduction

Molecular dynamics (MD) is a computational technique used in physics and chemistry to simulate the behaviour of atoms and molecules over time. It is particularly valuable for understanding the dynamic aspects of molecular systems at the atomic level to provide insights and comparisons to experimental observations. Simulations where all atoms are explicitly defined can quickly become too computationally expensive, especially when larger systems and interactions such as biomolecular systems, are simulated. Therefore, Martini [1] is a coarse-grain (CG) force field designed for simulating larger biomolecular systems at larger timescales. Unlike all-atomistic (AA) simulations that detail each atom and bond, coarse-grained models group several atoms into a single interaction site. This reduction in detail allows for larger time scales and systems to be explored, enabling researchers to observe phenomena that would be computationally prohibitive with all-atom models. The Martini 3 [2] force field extends its applicability to a wide range of biomolecules, including lipids, proteins, and nucleic acids. The force field parameters are optimized to capture essential interactions that emulate those of AA simulations. Proper parameterization is crucial for reproducing experimental observations and ensuring the reliability of simulation results. Deoxyribonucleic acid (DNA) is a fundamental biomolecule that carries the genetic instructions necessary for the development, functioning, and reproduction of all known living organisms. To be able to effectively simulate DNA in larger biological systems and observe their interactions over microsecond timescales, CG resolution must be used. Therefore, in this study, the martini CG forcefield parameters are optimised for DNA, with the focus on B-DNA. DNA exists in different structural forms, and among them, B-DNA and Z-DNA are two distinct conformations with unique characteristics. B-DNA represents the canonical right-handed helical structure that is commonly observed under physiological conditions. It is composed of a random sequence of A, T, G, and C residues. In contrast, Z-DNA is a left-handed helical conformation with a zigzag backbone geometry, composed of $(CG)_n$ or $(GC)_n$ sequences under high ionic salt concentrations [3]. B-DNA and Z-DNA are able to transition freely between the conformations upon a change in ion concentration [3]. Investigating these structural variations provides insights into the diverse roles that DNA plays within the cellular context. However, the parametrization of DNA residues in coarse-grain models, including Martini, poses significant challenges. This study aims to address these challenges and offer insights into the nuances of parametrizing nucleic acids.

2. Method

2.1. Computational Details

All simulations were performed with GROMACS 2021.5 [4].

2.2. All-Atomistic (AA)

All AA used the same computational parameters. The temperature of the production and equilibration simulation steps were controlled using the Bussi-Donadio-Parrinello thermostat (V-rescale) [5] and set to 300 K. Both the production and equilibration simulations used the Berendsen barostat [6] with isotropic pressure coupling. The reference pressure was set to 1 bar and the compressibility to 4.5e-5 bar^{-1}. Constraints for all-bonds were solved using LINCS [7] with a LINCS order of 4. The timestep of production run and equilibration runs were set to 2 fs. The production run was set to run for 100 ns. The electrostatic interactions were computed using smooth particle mesh Ewald (PME) [8] for the simulation in the NVT and NPT ensemble, and the production run. For the position relaxation and minimisation steps the cut-off at a distance 1 nm method was used. All mdp files that were used can be found in /grain/jodie/mdp-files.

2.2.1. Coarse-Grained (CG)

All CG simulations used the same computational parameters. The temperature of the production run and equilibration steps were controlled using V-rescale and set to 300 K. The equilibration step used the Berendsen barostat with isotropic pressure coupling. The compressibility was set to 3e-4 bar^{-1} with a reference pressure of 1 bar. For the production run the Parrinello-Rahman barostat was used with isotropic pressure coupling. The compressibility was set to 10e-4 bar^{-1} with a reference pressure of 1 bar. The length of the production run for CG was set to 2 µs with a timestep of 10 fs. The mdp files for martini simulations can be found in /grain/jodie/mdp-files.

2.3. Simulation Setups

2.3.1. All-Atomistic (AA)

All AA reference simulations were set up in identical ways. The AMBER OL21 [9] topologies required for simulations were downloaded from the “OL” Force Field Refinements for RNA and DNA simulations webpage [10]. The forcefield files used for the AA simulations can be found in the directory /grain/jodie/forcefields/amber14sb_OL21.ff. The topologies used for the DNA strands were generated using polyply [11] [12] , they can be found in the respective AA file for each molecule under the file

name “top”. The AMBER OL21 forcefield specific for polyply which can be found in /grain/jodie/polyply_1.0/polyply/data/ol21.

The completed structure files were placed in a dodecahedron simulation volume with a distance between the solute and box being 3 nm. A large distance between the solute and box is crucial for when the system is subsequently coarse-grained to prevent the larger beads from interacting with its periodic images.

The energy of the DNA structure was minimized first in a vacuum using the steepest-descent algorithm. It was then solvated with GROMACS solvate [4] using the spc216.gro water model. Ions were added to neutralize the system and to emulate biological conditions using GROMACS genion [4] with a concentration of sodium and chloride ions of 150 mM [13].

The system energy was minimized using the steepest-descent algorithm to remove the greatest strain that may be present and then the system was equilibrated in three steps. The first equilibration step is to allow the solvent molecules to adapt to the DNA structure using the md integrator. Generally, position restraints are used for this step, however, restraints were not necessary for the system and the solvent and DNA were allowed to adapt freely. The second equilibration step is where the system is brought up to the desired temperature of 300 K in an NVT ensemble. The third equilibration step is where the system is brought up to the desired pressure of 1 bar in an NPT ensemble. The AA simulations were run for 100 ns.

The final structure file produced from the production run is used as the starting structure for the coarse-grained simulation. First, the AA structure needs to be mapped to CG. This can be done using the github package fast_forward, downloaded from https://github.com/fgrunewald/fast_forward. Once downloaded and installed, the corresponding command ff_map can be used to convert AA structure and trajectory files into CG. The command allows for specific residues from the system to be mapped, thereby removing all other molecules or particles from the system. The mapping file that was used can be found in the directory /grain/jodie/forcefield/dna_cg_parameters/DNA/ for the double stranded and single stranded DNA molecules. The mapping files were obtained from Linus Grunewald and were based on previous RNA mapping parameters. For the isolated base molecules, the mapping file can be found in /grain/jodie/AA/dimers/ then under either adenine or thymine in the file /MAPPING/. The workflow for the AA simulations including all commands that were used can be found in the directory /grain/jodie/ called “commands.txt”.

2.3.2. Coarse-grained (CG)

The CG simulation steps were very similar to that of the AA. The necessary Martini 3 [2] topologies for coarse-grain simulations of the ions and water were downloaded from the Martini webpage [14] and can be found in the directory /grain/jodie/forcefields/. The mapped structure file was first minimized in a vacuum using the steepest-descent algorithm. The system was solvated with the martini water molecules from the martini water model using GROMACS solvate. The water file can be found at /grain/jodie/CG/W.gro. Ions were then added to neutralize the system and to attain biological conditions as was the case for the AA set up. Another minimization was done to remove the greatest strain in the system using the steepest-descent algorithm. The CG equilibration consists only of one step, where temperature and pressure is applied simultaneously, unlike the AA simulation steps, where that was done in two steps not including position relaxation equilibration step. The CG production runs were set to 2 us. The workflow for the CG simulations can be found in the file /grain/jodie/CG/commands.txt.

2.4. Source of Structure Files

2.4.1. B-DNA

B-DNA, as mentioned before, is characterized by its random mixture of residues. For simplicity, B-DNA with only A residues on one strand and only T residues on the other strand was used, composed of 15 bases. The relevant structure file was downloaded from http://www.scfbio-iitd.res.in/software/drugdesign/bdna.jsp. This structure file includes all hydrogen atoms on the residues. Before it can be used in simulations, the first and last residue names had to be changed as terminal residues do not have the exact same atoms as non-terminal residues. The name of the first residue of each strand was changed from DA (or DT) to DA5 (or DT5), likewise at the end of the strand the residue name was changed from DA (or DT) to DA3 (or DT3). These names correspond with the forcefield file used. To obtain single stranded DNA structure files, simply the irrelevant half of the structure file of the B-DNA double stranded was deleted.

2.4.2. Z-DNA.

The structure for Z-DNA was downloaded from the website http://web.x3dna.org/. This structure file does not come complete as the hydrogen atoms are missing. To add the hydrogen atoms the GROMACS gmx2pdb [4] command is used with the -ignh flag. The topology files that were generated using this command were discarded as polyply topologies are preferred. The command gmx2pdb happens to generate the hydrogen on the terminal 5' strand in the wrong order for the polyply topology, therefore the order of the hydrogen and the 5' oxygen were swapped around on each strand. Additionally, the

downloaded structure file also comes with a phosphate group on the 5' terminal of each strand that needs to be deleted. The first and last residues also need to be renamed to correspond with the terminal residues in the topology files, similarly as was done with B-DNA.

2.4.3. Adenine and Thymine

The structure files for the single bases adenine and thymine were created by removing the irrelevant atoms in the structure files from the B-DNA. The non-hydrogen atoms bound to the C1' carbon were all changed to hydrogen atoms.

2.5. Analysis

2.5.1. Parameterization

The parameterization of the CG force fields for DNA was done by comparing the bonds, angles and dihedrals of the AA mapped structure with the CG production run, as well as the distribution of those values. First the analysis was done on a single strand of A residues of 15 base pairs. To extract these interactions, the Github package fast_forward [15] can be used, as described earlier, with the command ff_inter. Then the graphs were plotting using a python script which can be found in /grain/jodie/scripts/. To ensure there are no strange artifacts in the graphs, the strand in the production run trajectory file was made whole using the command GROMACS trjconv with the flag -pbc whole. This ensures the molecule is not crossing the periodic boundary which can result in incorrect bond lengths.

Once the CG mapping is complete, enabling the accurate simulation of CG DNA requires an appropriate force field. Optimal force field parameters are determined by analysing the distributions of bonds, angles, and dihedrals in the AA-mapped trajectories. In this process, initial AA simulations of DNA are conducted, and the resulting trajectory files are mapped to CG. Subsequently, frequency distributions for bond lengths, angles, and dihedrals are generated. The values associated with the highest frequencies in each category are then selected for incorporation into the CG force field. The optimization process began with the utilization of a force field originally optimized for RNA simulations.

The force constants of the forcefield are altered based on the broadness of the distribution. If the distribution is too narrow, the force constant must be decreased. If the distribution is too broad, the force constant must be increased. Force constants determine the rigidity of the bond or angle that is defined, therefore by increasing or decreasing the force constant, the bond (or angle) will be allowed to attain less, or more values, respectively.

2.5.2. Clustering

Two bases of either adenine or thymine was placed in a small simulation box using the command GROMACS insert_molecules [4]. It is important to keep the periodic box small enough, so the bases have more frequent interactions and increase simulation speed by decreasing the amount of water molecules in the box. The vector sizes of the box were kept at around 3.8 nm. The following subsequent AA simulation, mapping to CG, and CG simulation are all identical to the method described previously. The analysis workflow for every step is shown in /grain/jodie/AA/dimers/commands.txt. Visualization states were edited in VMD, incorporating CG bonds using a custom Tcl script (cg_bonds-v2020.tcl). Radial distribution functions (RDF) were calculated using GROMACS rdf, and potential of mean force (PMF) profiles were generated with a python script /grain/jodie/scripts/df2pmf-compare.py. This script was obtained from /home/alex/PYTHON3/rdf2pmf.py and modified with the assistance of an AI language model based chatbot [16] to plot graphs from AA and CG together for better comparison. The PMF was calculated using equation 1, where the sFactor was given a value of -2.5, the variable 'small' is a constant of value $1.6 \cdot 10^{-6}$, and RDF values are read from the file generated by the GROMACS rdf command.

$$PMF = sFactor \cdot \ln(RDF + small)$$

Equation 1: PMF Formula.

The python script /grain/jodie/scripts/selectBoundFames2.py was used to select frames based on certain criteria, removing frames where the centre of mass between the two bases is greater than 10 Å. The clustering was done using the command GROMACS cluster with a cut off of 0.20 nm. This script was obtained from /grain/Alex/DNA/SHORTSTRANDS/A2/M3/BASE/NV3 and slightly modified to calculate the centre of mass of the base beads, rather than using the centre bead, with the assistance of an AI chatbot [16]. The overlap of the clusters was plotted using the python script /grain/jodie/scripts/ConformationalClustersOverlap.py. This script was obtained from /home/alex/PYTHON3/ConformationalClustersOverlap.py (no alteration was done).

2.6. Bonded interactions

In this section the bonded interactions that calculates the bonds, angles, and dihedrals in the forcefield file will be defined. The functions are obtained from the GROMACS 2023.3 documentation [4]. The bonds in the forcefield were calculated by representing the bond stretching as a harmonic potential, $V_b(r_{ij})$, Given by equation 2, where k_{ij}^b is the force constant in [kJ mol^{-1} nm^{-2}] and b_{ij} is the equilibrium bond distance in [nm], and r_{ij} is the bond distance in [nm].

$$V_b(r_{ij}) = \frac{1}{2} k_{ij}^b (r_{ij} - b_{ij})^2$$

Equation 2: Harmonic potential of two covalently bonded atoms, function type 1.

The harmonic force given by equation 3.

$$\boldsymbol{F}_i(\boldsymbol{r}_{ij}) = k_{ij}^b (r_{ij} - b_{ij}) \frac{\boldsymbol{r}_{ij}}{r_{ij}}$$

Equation 3: Harmonic force of two covalently bonded atoms, function type 1.

The angles in the forcefield were calculated using the restricted bending potential, $V_{ReB}(\theta_i)$, given by equation 4, which restricts the angle from attaining a value of 180 °. k_θ is the force constant in [kJ mol^{-1} rad^{-2}], and θ_i is the angle between three atoms (or beads) in degrees [°].

$$V_{ReB}(\theta_i) = \frac{1}{2} k_\theta \frac{(\cos\theta_i - \cos\theta_0)^2}{\sin^2\theta_i}$$

Equation 4: Restricted bending potential, function type 10.

Equation 5 gives the force of the restricted bending of the angle between three atoms (or beads).

$$F_{ReB}(\theta_i) = \frac{k_\theta}{\sin^4\theta_i} (\cos\theta_i - \cos\theta_0)(1 - \cos\theta_i \cos\theta_0) \frac{\partial \cos\theta_i}{\partial \overrightarrow{r_k}}$$

Equation 5: restricted bending force, function type 10.

The dihedrals in the forcefield were calculated using two different ways, depending on the atoms or beads. Some were calculated using the improper dihedrals, which are used to keep planar groups planar, given by equation 6. The angle between the planes of the atoms (or beads) ξ_{ijkl} is given in degrees [°], the force constant k_ξ is given in kJ mol^{-1} rad^{-2}.

$$V_{id}(\xi_{ijkl}) = \frac{1}{2} k_\xi (\xi_{ijkl} - \xi_0)^2$$

Equation 6: Improper dihedral potential, function type 2.

Others were calculated using the proper dihedrals, given in equation 7, where k_ϕ is the force constant given in kJ mol^{-1}, and the angle ϕ_s is given in degrees [°].

$$V_d(\phi_{ijkl}) = k_\phi (1 + \cos(n\phi - \phi_s))$$

Equation 7: Proper dihedral potential, function type 9.

3. Result and Discussion

3.1. Mapping and Parameterization

In this section the mapping of all-atomistic (AA) DNA to coarse-grain (CG) DNA will be discussed. Mapping is where a number of atoms are grouped into a single bead to decrease the resolution of the structure. The interactions of these beads are then parameterised to emulate the AA model as accurately as possible. The original mapping files and forcefield parameters were taken from Linus Grunewald, who previously worked on and parameterised RNA. Only the forcefield file of RNA is parameterised for DNA, as the mappings for DNA were already provided by Linus.

Figure 1 shows how an adenine (A) single strand of a DNA molecule is mapped to CG (thymine (T) single strand is given in the appendix, A 1.). The figure shows the terminal and nonterminal residues, as the atoms contained in each is slightly different and therefore mapped differently. The length of the strands in the structure files used is 15 bases. In the backbone of the 5' terminal end of a DNA strand (labelled DA5 for adenine in figure 1), the BB3 bead contains an extra alcohol group, while in the 3' terminal end group, the BB2 bead contains an extra hydrogen. The backbone mapping remains the same for all residues. The BVS bead is a virtual site i.e. its position is defined by the position of BB1, BB2, and BB3 and has no direct bonds to the other beads. In the A residue, there are two virtual sites viz. SC3 and SC6. SC3 is defined by SC5, SC2, and SC4, while SC6 is defined by SC2 and SC5. In the T residue, there is only one virtual site, SC3, which is defined by SC2 and SC4. This mapping allows for a structure that resembles DNA, where the bases are planar, and the backbone is flexible enough to be able to curve in a helical fashion.

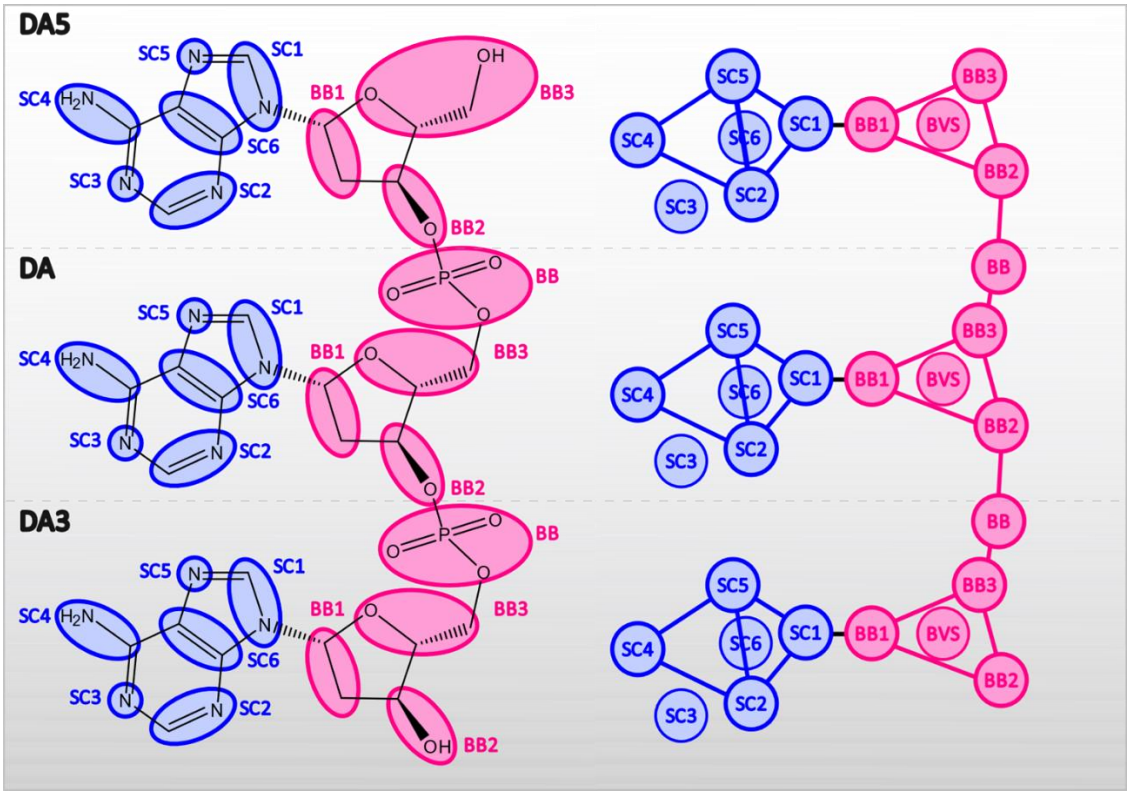


Figure 1: Atomistic to coarse-grain mapping of A residues (left). 5 indicates terminal 5' end, and 3 indicates terminal 3' end.

Figure 2 shows the distributions of the AA trajectory of the A residue in full lines. The top graphs show the distributions of solely the backbone beads (BB) interactions, while the bottom graphs include backbone with sidechain (SC) interactions, or solely sidechains. In this case there is no differentiation made between terminal residues and nonterminal residues, all the values of the bonds, angles, and dihedrals of the same beads are plotted in the same graph. Perhaps for more accuracy, differentiation between these residues can be made, as there are some differences in forcefield parameters. The dashed lines in figure 2 represent the CG trajectories after a production run with the optimised forcefield. The plot with the same colours corresponds to the same group of beads that are being analysed. The legend describes which group of beads the value is taken from.

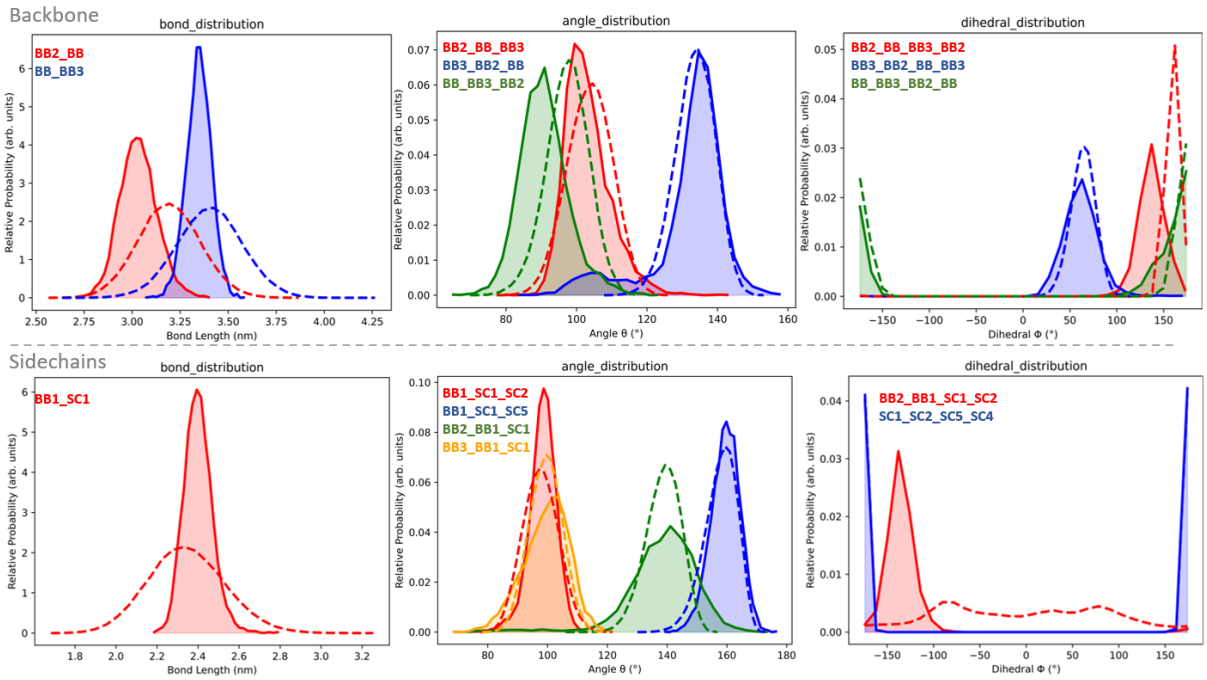


Figure 2: Bond, angle, and dihedral distribution for adenine single strand. The top figures are solely backbone distributions. The bottom figures include side chains. The dashed line is the CG distribution. The full line is the AA mapped distribution.

The distribution overlap of the bonds is not the most optimal. The value of the most frequent bond length is reasonably similar to that of AA. The most obvious difference is the height and width of the distribution. AA bonds tend to have a very narrow distribution, which indicates that these bonds take on a smaller range of values, corresponding to higher force constants (K). The CG distribution is very broad, indicating that the corresponding force constants in the force field are too low, and the bonds can assume too many different lengths. Many attempts were made to increase this force constant to allow for a narrower distribution, however, even minor tweaking rendered the simulation unstable. For example, changing the force constant for DA BB1_SC1 from 5000 (table 1) to 6000 caused system instability within a very short time. The exact cause of this is unknown. To pinpoint the cause, more understanding of the cause and effect of changing these parameters must be known. It may simply be that CG simulations require more bond flexibility.

The overlap of the angle distributions is relatively good. Most of the values for the most frequent angle match those of the AA model very well, except for the angle BB_BB3_BB2. The value for the angle in the force field was 89°, however, it seems that there is some other strain in the molecule that is preventing this angle from reaching the required value. Changing this angle to even smaller values to ‘force’ it to the correct value did not seem to make it budge. From looking at the visualisation of the trajectory in VMD, it wasn’t obvious why this is the case.

The dihedral distributions also show significant overlap. The only dihedral which doesn’t match exactly is the dihedral BB2_BB_BB3_BB2 in the backbone. Potentially some kind of strain in the

structure was preventing this dihedral from attaining the optimal value. More analysis needs to be done on the dependency of other angles and bonds on the size of the dihedrals to get a better understanding. The dihedral defined as BB2_BB1_SC1_SC2 caused instability in the CG simulation. For short simulation times up to 250 ns there were no issues, however, at longer simulation times up to 2 us, the simulation crashed near the end. Once the dihedral was deleted the simulation ran without issues. The CG dihedral distribution of BB2_BB1_SC1_SC2 in figure 2 (bottom right) shows that this dihedral without restrictions exists as essentially all angle values, with less frequency near 180 °. This dihedral determines the angle of the plane on which the base resides with respect to the backbone. In the atomistic model, the bases lie nearly completely parallel on top of one another (figure 5, middle left), in CG, the bases tend to twist and turn in all different directions (figure 5, left). It could be that the angle is only favourable when the DNA is in its compact helical structure where base stacking is prominent and outside of this conformation, the angle causes too much strain as the bases interact with water and are forced into random positions by the backbone if the backbone can attain any random conformation. The final bonded parameters are given in the appendix A 2.

The bonds, angles, and dihedrals of the T-strand were also analysed. This is given in figure 3. Again, in general there is reasonable overlap. The major difference is again the height and width of the bond distributions, but also the BB2_BB bond length. Although in the forcefield this bond length is defined as the value that corresponds with the AA bond length, something in the structure is preventing this bond from becoming any smaller. In the process of parameterising the T-strand, the simulation very quickly became unstable when angles and dihedral values were changed to match that of the atomistic mapped distributions when the dihedral BB2_BB1_SC1_SC2 was still defined. This was unlike the A-strand where it took quite some time before the simulation became unstable. Once this dihedral was removed, the simulation was stable and was able to go to completion. The T-strands force constants were not changed upon the removal of the unstable dihedral to better match the required distribution. The final parameterisation values for the T-strand are also given in the appendix A 2.

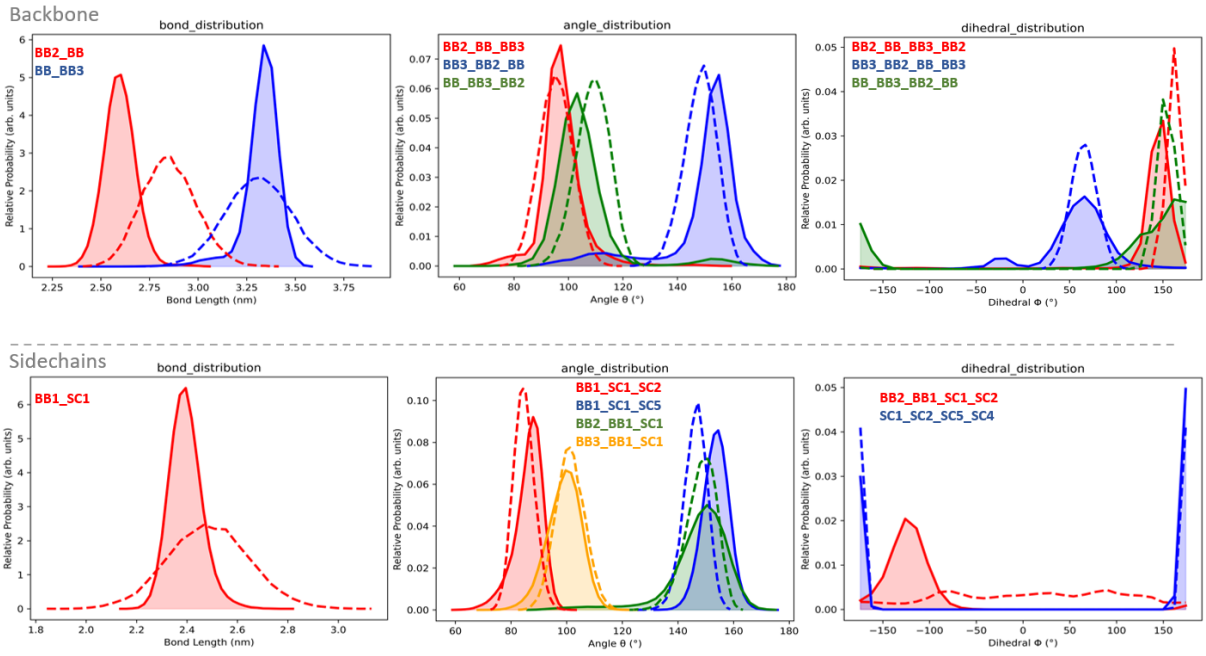


Figure 3: Bond, angle, and dihedral distribution for the thymine single strand. The top figures are solely backbone distributions. The bottom figures include side chains. The dashed line is the CG distribution. The full line is the AA mapped distribution.

It is interesting to note that the BB2_BB bond length for the T-strand is significantly smaller than the A-strand. This is shown more clearly in figure 4, where T-strands have dashed lines and A-strands have full lines. For an extra comparison of the differences in backbones of various strands, the GC alternating strand of Z-DNA was also added in dotted lines. One colour represents one group of beads that need to be compared. There are some major differences in the parameters of the backbones. The main difference between A-strand and T-strand is the bond distribution of BB2_BB, which ranges from 2.4 nm to 3.0 nm, and the angle distribution of BB3_BB2_BB, which also has a large

discrepancy. This seems to indicate that there is no “one-size-fits-all” approach when it comes to parameterising backbones even though they all contain the same beads with the same atoms.

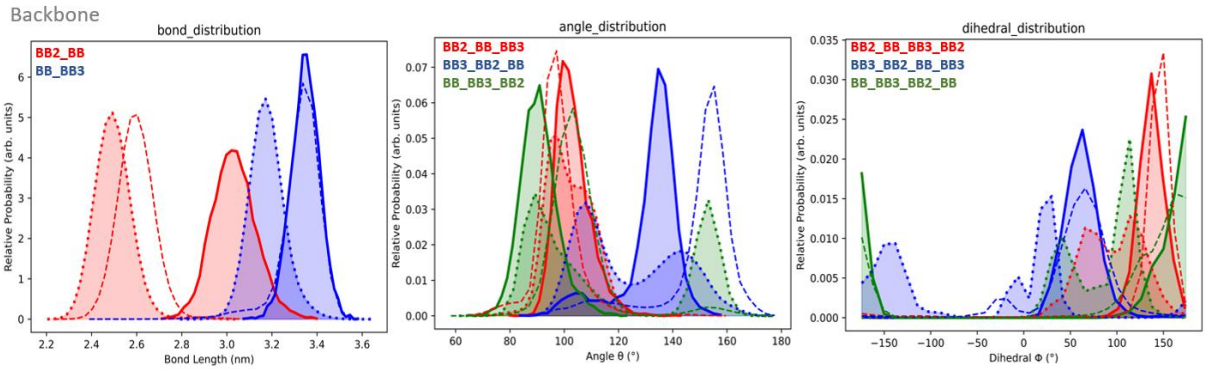


Figure 4: Distributions comparing A-strands (full lines) with T-strands (dashed lines) and alternating GC residues in z-DNA (dotted lines) in the backbone of AA mapped trajectories.

The dihedral distribution for BB_BB3_BB2_BB, the Z-DNA strand has two peaks quite far apart, one below 50 ° and one above 100 °, while the A and T strand are both above 150 °. Z-DNA is much more prone to having two peaks for angle and dihedral distributions. This is to be expected as DNA is known for its ‘zigzag’ backbone structure due to the alternation of purine and pyrimidine $(GC)_n$ or $(CG)_n$ where the nucleotides are in different conformations [17]. B-DNA can also have a $(GC)_n$ or $(CG)_n$ repeating sequences where this is not the case. Previous MD simulations found using the GROMOS96 forcefield that the transition from B-DNA conformation to Z-DNA occurs when the ionic strength of the solvent is changed [17]. Z-DNA conformation is favourable over B-DNA in strong ionic conditions. That is because the Z-DNA conformation is more compact and therefore has more repulsion of the negative charges in the backbone. When the solvent has a large concentration of ions, these ions can shield some of those charges, decreasing the net repulsion. When ionic concentration decreases, the backbone will have too much strain caused by the negative charges and the conformation will transition into B-DNA, which is the less compact structure [17]. This shows that explicit charges and electrostatic interactions of the backbone with the solvent (and itself) play a crucial role in the structure of DNA. To successfully model Z-DNA and B-DNA transitions in CG, it is important to be able capture these electrostatic interactions in the backbone beads.

Overall, the parameterisation for bonds, angles, and dihedrals was mostly successful aside from the force constants. In general there was significant overlap of the distributions with a few variations. Even though the parameters seem to match quite well, the CG single strand DNA, when simulated, does not attain the base-stacking helical structure that DNA is known for (figure 5), for neither the T-stand nor A-strand. The ideal structure of a single strand DNA for the A-strand and T-strand is shown on the middle left and on the right, respectively. This is the AA mapped to CG structure.

The CG A-strand shown on the left, shows significant elongation of the backbone where the residues are nowhere near close to one another. Throughout the whole trajectory the structure is completely random. No base stacking or favourable base interaction in general is observed (trajectory can be viewed at /grain/jodie/CG/ss-a15/EQ/viz.vmd). The exact same is observed for the T-strand (middle right). One minor observation is that the bases are slightly more aligned in the correct orientation, as opposed to the A-strand where the bases point in random directions. The T-strand, although more aligned, shows next to no base stacking throughout the trajectory (trajectory can be viewed at /grain/jodie/CG/ss-t15/EQ/viz.vmd). The parameterised forcefield was applied to the double stranded DNA with one T strand and one A strand. The result is shown in figure 6 (right). Interestingly, the molecule shows significant base stacking, albeit with intercalation of AT bases and not in the way that resembles the double helical structure of DNA. What is interesting is that the stacking occurs and remains stable throughout the 2 μs simulation time, and that the stacking has some rotation. This shows that there is a propensity for the coarse grain bases to stack in some helical orientation. From simple observation, it seems that in the single strand case, there is far too much space being allowed between the bases, preventing any long lasting favourable interactions between them. Another aspect to consider is that the backbone can also contribute to keeping the bases away from one another, rather than keeping them in close proximity. In the study conducted by Chi H. Mak [18], it was revealed that base stacking and backbone conformational strains have opposing effects [18]. The research, which utilized large-scale computer simulations, demonstrated that the measured base stacking free energy in a DNA complex is relatively mild. This mild energy is attributed to the counteracting entropic strain of the backbone with the favourable entropic interaction of base stacking [18]. The paper will be discussed in more detail in the next section. Figure 6 of the double stranded CG DNA simulation shows that if the backbone strain is removed, base stacking becomes favourable. This shows that in the CG model, the favourable stacking interactions between bases are not strong enough to overcome the unfavourable backbone conformational strain, whereas in the AA model, the base interactions are strong enough. To gain more insight into the difference in base interactions of the CG and AA model, a clustering analysis was done. This will be described in the next section.

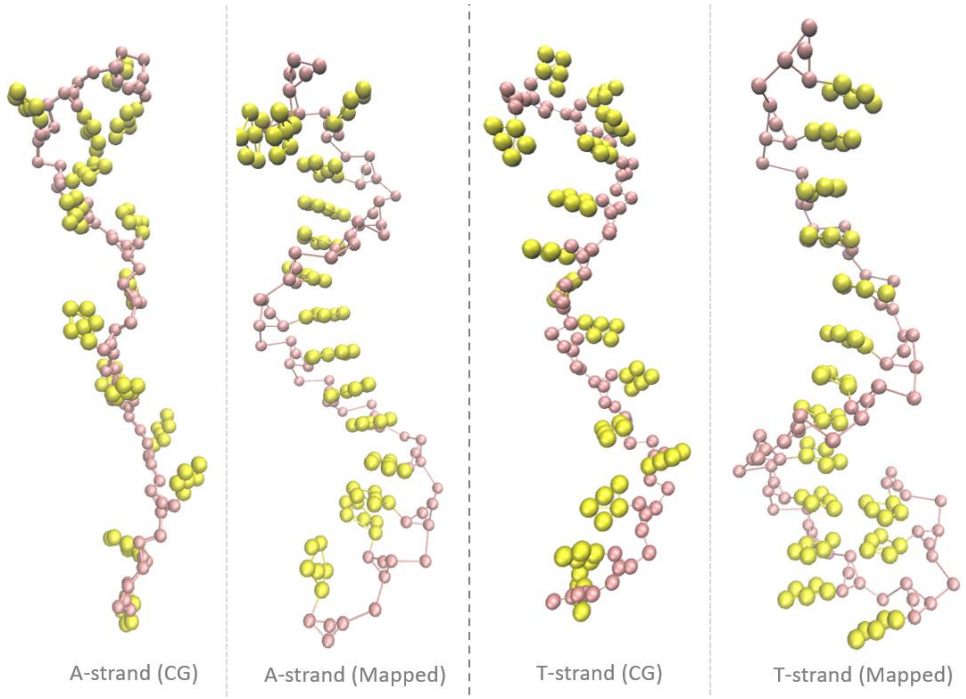


Figure 5: Comparison of A-strand structure CG (left) to AA mapped (middle left). Comparison of T-strand structure CG (middle right) to AA mapped (right). Strands are composed of 15 bases.

Another aspect to note in the double stranded CG simulation is that there is no base pairing interaction seen at all, even though the beads have corresponding accepter/doner type interactions between them. In the atomistic model, the single strands can be seen base pairing with itself. This is shown for the T-strand in figure 5 (right). This would also occur for the A-strand.

According to a study that was done on the driving forces of DNA base pairing using calorimetry by Dragan et al. it was found that the contribution of hydrogen bond pairing between the bases is purely entropic [19]. There is essentially no negative enthalpy driving the hydrogen bonding between the bases [19]. The increase in entropy comes from the removal of water that is hydrogen bonded to the bases and releasing it back into the bulk solvent, this alone is enough to drive the pairing of strands [19]. Therefore, to better drive base pairing in CG simulations, instead of focusing on strengthening base-pair interactions, the focus should be placed on base-water interactions, so that there is a large entropic gain from the bases pairing with one another. For example, the number of hydrogen bonds between water and bases should be analysed as that gives an indication of the entropic gain that is experienced by releasing those waters in the bulk solvent. To model a similar entropic driving force in CG, a similar number of particles should also be released from 'bound' states. Also, if the hydrogen bond of CG nucleic acid residue and water molecules are not as strong as in the AA model, there is even less entropic gain, as the water movement is already less restricted. The effect of maximising entropic driving forces in CG will be discussed in section 3.3.

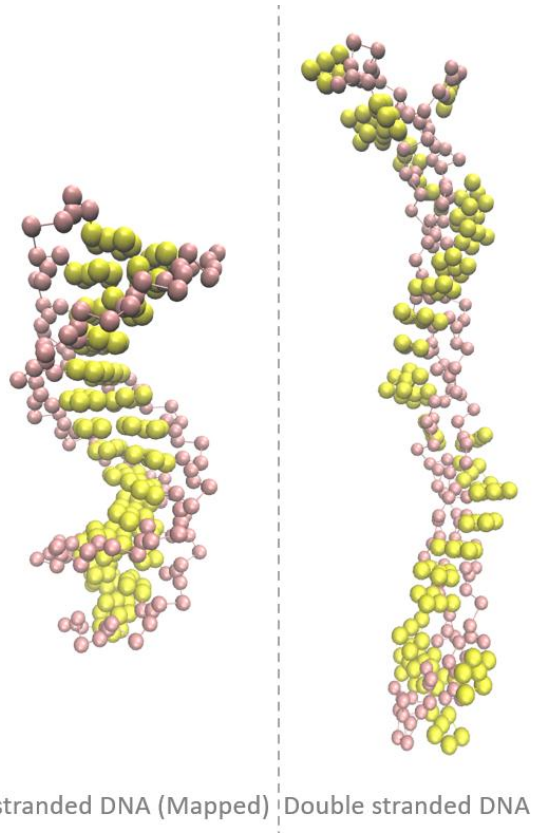


Figure 6: CG DNA structure comparisons of length 15 bases. Double stranded DNA with only A residues on one strand and only T residues on the other strand mapped from AA (left). Double stranded DNA from CG production simulation (right).

3.2. Clustering Analysis

In this section the clustering of bases will be discussed. In this analysis, two bases (AA or AT) were simulated in a solution of water and ions of concentration 0.15 M. Clustering is useful to determine the most probable conformation of a selection of molecules, which gives insights into their interactions. The clustering analysis was done to compare the base stacking interactions of AA with CG to develop a better understanding of base-base interactions are in the CG model. As mentioned before base stacking interactions must be able to outcompete the backbone strain of the DNA molecule. Therefore this analysis will shed light on the relative strength of the base-base interactions and determine whether the CG base-base interactions compare with those of AA.

The analysis was done twice using different beads. First, the bases were mapped using the same bead types as were used in the CG DNA simulation, this will be referred to as “beads 1”. Afterwards the beads were changed to incorporate more attractive forces with like beads, termed “beads 2” (the exact same was done for thymine). To add attractive forces, the label ‘h’ [2] was added to the bead type. Adding the ‘h’ label will increase self-interaction with like beads with the same label

[2]. For example, interaction of C with C beads will be increased if they both have the 'h' label, but interactions of C with N beads will have no effect even if they both have 'h' labels [2]. Table 1 shows the bead types for adenine and thymine.

Table 1: Bead types of adenine (left) and thymine (right). Beads 1 refers to the original beads used. Beads 2 refers to beads with added attractive forces.

		Adenine		Thymine	
		Beads 1	Beads 2	Beads 1	Beads 2
1	BVS	TC4	TC4	TC4	TC4
2	SC1	TN1	TN1h	TN2	TN2h
3	SC2	TN3a	TN3ah	TP1a	TP1ah
4	SC3	TN5a	TN5a	TN5d	TN5d
5	SC4	TP1d	TP1dh	TP1a	TP1ah
6	SC5	TN3a	TN3ah	TC2	TC2h
7	SC6	TC6	TC6	--	--

Figure 7 shows the result of changing bead types for adenine. The graph on the left shows the frequency of specific orientations at which the bases interact. The x-axis refers to the different orientations, named 'clusters'. Cluster id 1 shows the most common orientation found in the simulation. Note that frames in the simulation where the centres of mass between the bases are greater than 10 Å have been removed. Cluster id 2 shows the second most common orientation, and so on. The frequency of each orientation is compared between AA and CG to see if there are any major differences in how the bases interact with each other. For example, if the AA model shows a high frequency of stacking in one orientation, ideally the CG model will show a similar amount. Figure 7 shows that this is mostly true for Beads 1. There seems to be significant similarities in the frequency of the different orientations or "clusters". The nature of the first four most common clusters is shown in figure 8.

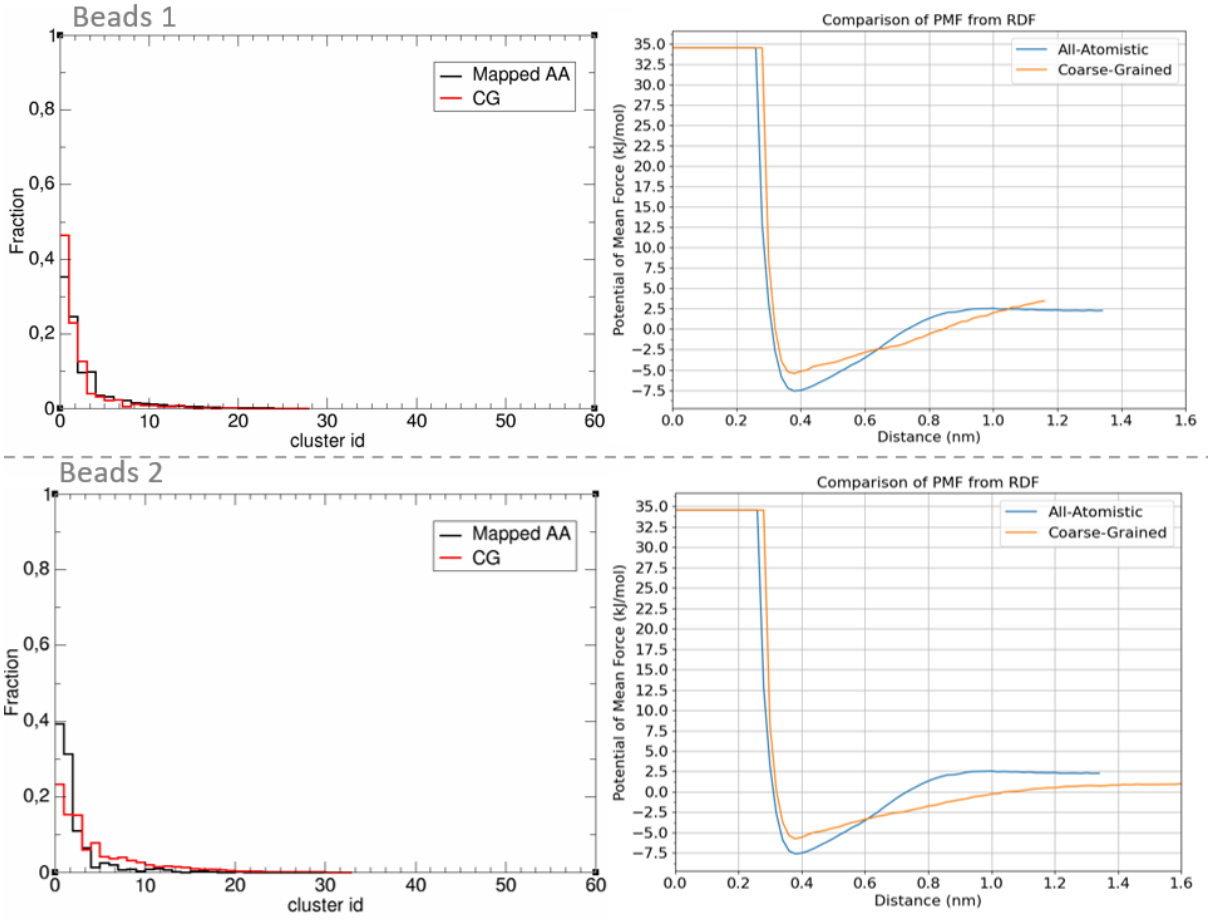


Figure 7: Clustering analysis of adenine dimer. Frequency of cluster orientations (left graphs) and potential of mean force versus distance (right graphs). Beads 1 have the same beads as in the DNA simulated prior (top row). Beads 2 have extra attractive forces for like beads (bottom row).

The graphs on the right (figure 7) describe the potential of mean force as a function of distance between the centres of mass. The atomistic model seems to have stronger attractive interactions, which is indicated by the depth of the potential well. The depth of the well of atomistic model, viz. -7.5 kJ mol^{-1}, seems to be fairly larger than that calculated by literature, which generally has values above -4 kJ mol^{-1} for any of the stacked bases AA, CC, GG, and AT [18]. In this study, computer simulations of the bases were done in a solvent box of 10 000 solvent molecules using constant NPT Monte Carlo [18]. The 'stacking' of bases was also more selective than in this study, as only expected stacking geometry of the bases were chosen for analysis [18], whereas in this study a cut-off of 10 Å was used from the centres of mass.

The CG model seems to have an almost linear relationship between potential of mean force and distance, which is not the case for the AA model. The repulsive forces for the AA model seem to taper off at around 2.5 kJ mol^{-1}. For the CG model, the repulsive forces continue to increase almost linearly, surpassing that of the AA model. The large repulsive forces are likely due to the simulation box being too small. The simulation box length should be increased to at least 3.8 nm to capture the

forces at longer distances and to determine at what point the CG repulsion forces plateau. The graph shows that it is important for the CG bases not to have a distance greater than 0.8 nm, or they start experiencing repulsive forces. It also shows that below going below 0.8 nm, the rate at which the CG bases experience more attractive forces are far lower compared to AA. In saying that, it is interesting to note that the frequency of clusters match quite well with the atomistic model.

For beads 2, more attractive forces for like beads were introduced. Interestingly, the depth of the potential well only seemed to increase a very small amount. Even more interestingly, the bases don't seem to interact in a stacking like orientation as well as with beads 1 (less attractive beads). This can be observed in the vmd visualisation at /grain/jodie/CG/dimers/adenine/beads1/Analysis/viz.vmd and /grain/jodie/CG/dimers/adenine/beads2/Analysis/viz.vmd. The frequency of clusters for the CG bases seems to have a much flatter distribution than with beads 1. This indicates that there is less affinity for particular orientations than others, that is, the first few conformations are only slightly more favourable than the others, as opposed to far more favourable. This aligns with the potential of mean force plot which shows that the potential well is much less steep. The repulsive forces don't start until well up to a distance of 1 nm and the rate of change of the potential force between the minimum and 1 nm is significantly smaller. This means that conformations at increasing distances are relatively (to AA) too favourable past around 0.6 nm (and not favourable enough below that value). Another difference is that the new beads have far less repulsive forces, which is expected.

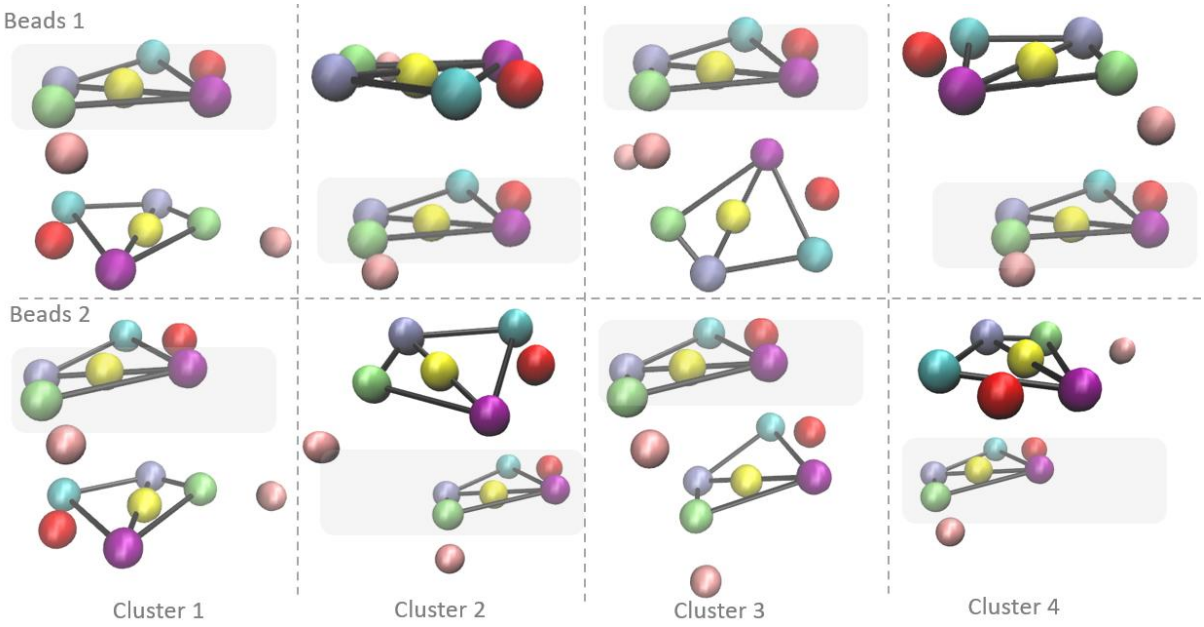


Figure 8: visualisation of the first four clusters of adenine. The same residue is highlighted in grey and kept at the same orientation for clarity.

Figure 8 gives the comparison of the beads 1 and 2 visualisation of the first four most common cluster orientations of adenine. In both cases, the orientations are not as parallel as one would expect, and they don't seem to favour any particular direction in how the individual beads align with the beads in the opposing base.

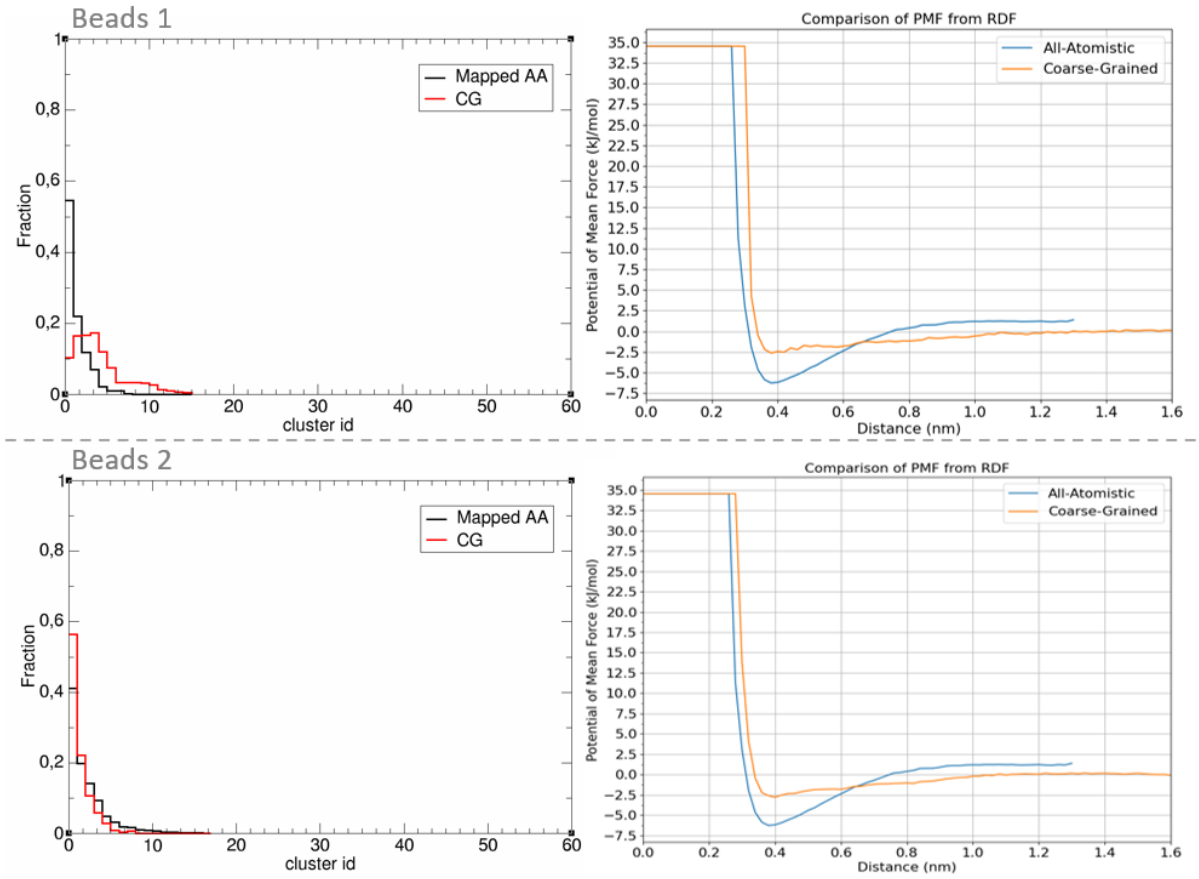


Figure 9: Clustering analysis of thymine dimer. Frequency of cluster orientations (left graphs) and potential of mean force versus distance (right graphs). Beads 1 have the same beads as in the DNA simulated prior (top row). Beads 2 have extra attractive forces for like beads (bottom row).

The same analysis was done for thymine. Figure 9 shows the frequency of orientations of CG compared to AA, and the potential of mean force versus distance. Somehow for thymine, the complete opposite is true; the potential of mean force shows similar patterns as that of adenine for beads 1 and beads 2, however, the frequency of cluster orientations is the opposite of what we expect based on the previous explanation. More research and analysis needs to be done to explain this phenomenon. Figure 10 provides the visualisation of the four most common orientations. Neither trajectories of beads 1 nor beads 2 show general base stacking behaviour like adenine beads 1 did. The trajectories can be seen at /grain/jodie/CG/dimers/thymine/beads1/Analysis/viz.vmd or /grain/jodie/CG/dimers/thymine/beads2/Analysis/viz.vmd. In general, the thymine bases seem to be

much further apart from one another compared to adenine and their orientations are much less parallel. Interestingly, the frequency of clusters of beads 2 of thymine seems to match quite well with the AA.

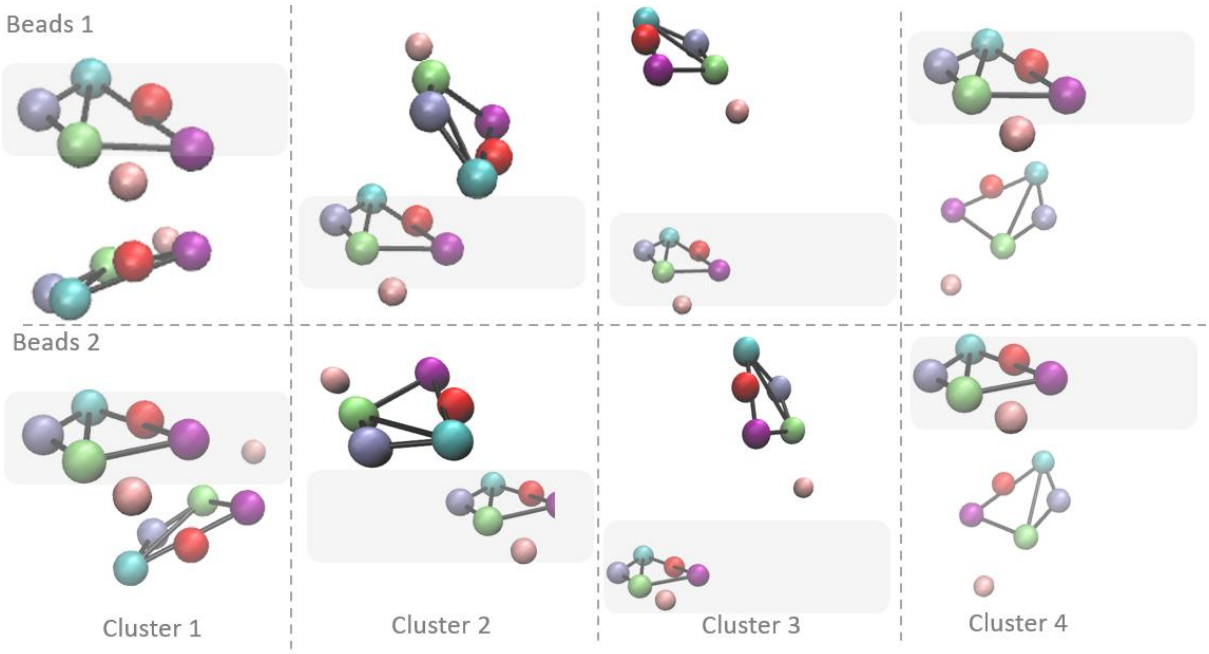


Figure 10: Visualisation of the first four clusters of thymine. The same residue is highlighted and kept at the same orientation for clarity.

The clustering analysis wasn't entirely successful in pinpointing the nature of base stacking behaviour. It did provide some useful insights such as at what point the bases start feeling repulsive forces and that the AA bases seem to have a much stronger attractive force. There seemed to be little correlation between the PMF graph and the nature of the clusters. Adding more attractive forces between like beads did not seem to impact the depth of the potential well enough to match that of AA. It could be the case that due to this, the strain of the backbone cannot be overcome. Additionally, the BB1 beads in the backbones are far enough apart (>0.8 nm) where the bases are only feeling repulsive forces rather than attractive forces. This may explain the propensity of the bases to reside in all different directions.

This analysis mainly focused on the π-π interactions of the bases with enthalpic driving forces as a focus point. According to a study done on base stacking driving forces, enthalpic driving forces is only secondary to entropic driving forces [18]. It was found that by far, the main driving force in keeping the bases together, is "nonhydrophobic solvent entropy", where the solvent tries to minimise the interaction it has with the solute [18]. The primary conclusion from the study states that the entropic gain from base pairing must be great enough to compensate for the entropic loss from backbone conforming [18]. This explains why the enthalpic focused approach to parameterising DNA hasn't

proven fruitful thus far. For future parameterisations of DNA, the interactions between water and the bases need to be carefully emulated from AA. Currently, the interactions between water and bases are too favourable so the water does not form a significant crystal like conformation around the DNA and creating a local decrease of entropy. If this were the case, the environment would react by trying to minimise the decrease in entropy by keeping the bases as close together as possible, forcing the backbone to compact. Figure 6 shows that, without the opposing force of the backbone, that the entropic force is great enough for base stacking to occur.

3.3. Maximising entropic driving forces

In the previous section it was found that the structure of DNA is predominantly influenced by entropic driving forces according to the literature [18] [19]. That is, the compact helical structure of DNA caused by base stacking, and the attachment of two single strands together, are both driven by expelling 'bound' water back into the bulk solvent. This section aims to maximise the entropic driving forces of the CG adenine single strand simulation by altering the bead types.

The first consideration was base stacking. Base stacking is facilitated by minimizing the interaction of water with the bases, a phenomenon known as the hydrophobic effect. In this context, the hydrophobic effect refers to the tendency of nonpolar molecules, such as the hydrophobic regions of DNA bases, to aggregate and exclude water molecules, driving the stabilization of the stacked conformation. To emulate this in CG, the bead types of the base that do not hydrogen bond were changed to very apolar beads. SC6 and SC1 were changed to the most apolar beads possible, viz. type C1, because these beads are in the centre of the base (see figure 1, section 3.1) and this would penalise the water for residing in between the bases. The attractive to like beads label 'h' was added just to provide more incentive for the bases to stack closely and penalise the interaction with water even more. SC2 and SC5 were changed to C4. This is less apolar than C1 (the scale for apolar beads ranges from 1-6 [2]) because SC2 and SC5 are on the edges of the base that will always have some kind of interaction with water, unlike SC6. Therefore, the hydrophobic nature of those beads shouldn't be too high or the bases might pair sideways with each other to minimise contact with water. The interactions with water also shouldn't be too favourable or the bases may be too easily solvated. The exact nature of these interactions in CG aren't entirely understood for DNA, but this is hypothesized to be the case. These beads were also given 'h' labels to encourage stacking.

SC3 and SC4 are the beads that are composed of the atoms from AA that partake in base pairing. Base pairing was found to be purely driven by entropic driving forces and has little to do with the strength of the hydrogen bonds. The entropic gain arises from water molecules, which become immobilized through hydrogen bonding to the base, being subsequently released into the bulk solvent.

Base pairing is also a very subtle interaction, in CG, the issue is often that pairing does not occur, and bases simply slide past one another, intercalating. To mitigate this, SC3 and SC4 were modified to have maximum polarity (P6). This modification appeared effective in preventing base intercalation. The polar nature of the base tip is hypothesized to face significant penalties when transported between the substantial apolar regions in the middle of the base. SC3 and SC4 also have the accepter and doner hydrogen bonding label, respectively, that would correspond to the atomistic structure (figure 1).

The backbone had minor changes. First BB3 the reduced attractive forces label ‘r’ was removed, to help minimise backbone conformational strain. BB1 the ‘h’ label was added for the same reason. The virtual site BVS was made entirely apolar, denoted as C1 with an attractive label 'h.' However, the impact of this modification on the conformation of the backbone remains uncertain, as the optimization of this bead type is not yet fully realized. The complete bead types for the adenosine residue of DNA are given in table 2.

To maximise the entropic gain of water molecules being expelled, the CG water beads were changed to tiny waters (TW) as this would increase the number of particles that are being released. Changing the water molecules to polarised beads could also increase this driving force, as more water would be attracted to the polar beads.

Table 2: Adenosine residue. The DA column were the original bead types, DA2 are the new bead types.

		Adenine	
Number	Bead label	DA bead type	DA2 bead type
1	BB	SQ5	SQ5
2	BB1	TC3	TC3h
3	BB2	TP1	TP1
4	BB3	SN4ar	SN4
5	BVS	TC4	TC1h
6	SC1	TN1	TC1h
7	SC2	TN3a	TC4h
8	SC3	TN5a	TP6a
9	SC4	TP1d	TP6d
10	SC5	TN3a	TC4h
11	SC6	TC6	TC1h

The result of the modification of the bead types is shown in figure 11 (left). This snapshot is from a random part of the CG trajectory about halfway through the 2 μs simulation time. The CG

structure shows significant amount of base stacking interactions, and it is clear that the penalty of the backbone strain is overcome by the favourable stacking interactions, and the backbone has conformed into a helical structure. This was not observed in prior simulations at all. What can be observed from the trajectory (/grain/jodie/CG/ss-a15/EntropyMax/EQ2/viz-run.vmd), is that the structure is not as stable as the AA simulation. The conformation of the nucleic acid is still too dynamic, and the bases seem to have a difficult time remaining neatly stacked on top of one another. It is hypothesized that this is still due to the opposing strain of the backbone. More research should be done in attempting to lower the strain of the backbone to see if this helps the structure to remain stable. The balance of the two forces have not yet been optimised. From the literature described previously by Mak, Chi H. the two opposing forces should cancel out in such a way that the base pairing free energy is only about 1 kcal mol^{-1} [18]. The free energy of the base stacking in the CG single strand should be analysed to see if it meets this value. In future research, other attempts can be made at lowering the backbone strain by, for example, removing restrictions for the backbone in the forcefield.

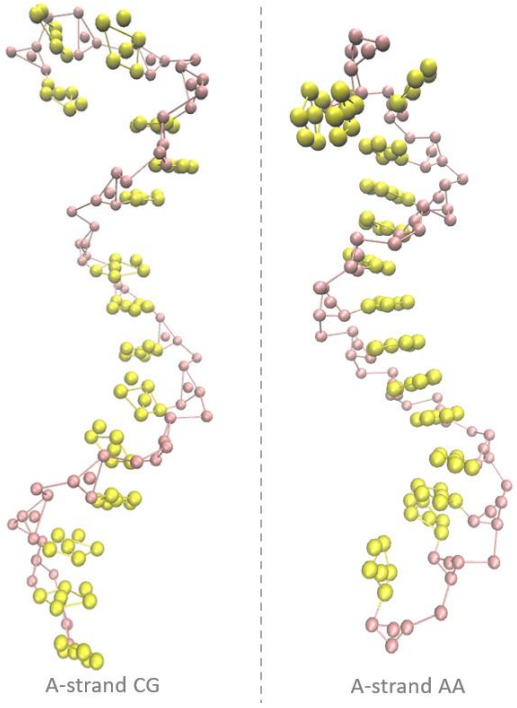


Figure 11: Comparison of the CG adenine single strand with new bead types (left) to the mapped adenine single strand structure from AA (right).

Base pairing is the other major aspect of the structure of DNA. Figure 12 shows a stable conformation found in the last half of the trajectory (/grain/jodie/CG/ss-a15/EntropyMax/EQ2/viz-run.vmd) where the single strand has started base pairing with itself. It can be seen that the bases stay pointed towards each other and do not start intercalating as is seen with previous simulations. For example, the simulation found at /grain/jodie/CG/ss-a15/EntropyMax/EQ1/viz-run.vmd shows

significant propensity of the bases to start intercalating once they get close to on another. In this simulation, the beads were not as relatively polar or hydrophobic as the new beads. This indicates that having very polar beads at the front (SC3 and SC4) and very hydrophobic beads in the middle (SC6, SC2, SC5, and SC1), stops the bases from intercalating when they are in base pairing conformation. That the single strand base pairs with itself corresponds to all-atomistic simulations as AA simulations of single strands also base pairs with itself, but the conformation of the AA single strand is much more stable and the backbone does not cross over itself in a loop like in figure 12. For future research it would be interesting to see if base pairing still occurs with normal sized waters (ideally normal size water is used to decrease computational costs). It may help to polarise the waters to increase their interaction with SC3 and SC4 polarized beads.

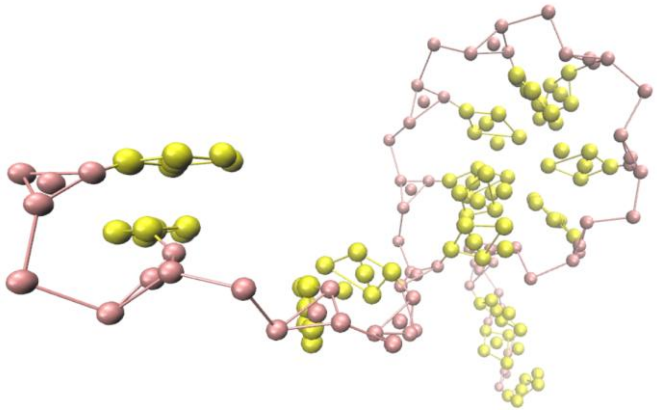


Figure 12: Observed base pairing of the adenine single strand in CG simulation with new bead types.

Conclusion and outlook

This study focused on parameterising the Martini 3 force field for coarse-grained B-DNA, specifically for adenine and thymine residues. The parameterization process involved utilizing all-atomistic (AA) mapped trajectories obtained from B-DNA simulations conducted with GROMACS 2021.5. The distributions of the bonds, angles, and dihedrals of AA against CG showed considerable overlap. However, The CG bonds were not able to attain the narrow distribution of the AA distributions due to any alteration of the force constants causing system instability.

Simulations of single strands of A or T residues using the optimized force field revealed a loss of compact helical structure and base stacking orientations. When comparing base-base interactions in a solution, the AA model exhibited slightly stronger interactions (-7.5 kJ/mol) compared to the CG model (-5 kJ/mol) in a clustering analysis with two single bases (AA or TT).

Interestingly, the clustering analysis didn't provide clear insights into the absence of observed interactions. According to a study using computer simulations on the driving forces on base stacking by Chi H. Mak, base stacking is largely a result of entropic driving forces, where the solvent molecules

aim to minimise contact with the solute [18], which counteracts the opposing force of the backbone strain. Simulations of a double-stranded DNA molecule with optimized parameters showed significant base stacking between bases from alternating strands where the backbone is significantly elongated, indicating that base stacking becomes favourable when backbone strain is not a significant factor. Or in other words, when the entropic penalty of the backbone strained conformation is less than the entropic gain of base stacking to minimise solvent interaction, base stacking is prominent and stable throughout the simulation. Increasing the enthalpic driving force by making the base have more favourable interactions did not compensate for this. What wasn't observed in the simulation was base pairing, even though the beads of the residue contained corresponding acceptor/donor character. Another study by A. I. Dragan et al. showed using calorimetry measurements that base pairing is in fact purely driven by entropic forces, and that enthalpic contributions to base pairing is negligible [19].

A final attempt in this study was made to change the bead types of the CG single strand DNA to maximise the entropic driving forces as was described in the literature. The beads residing in the middle of the base (SC6 and SC1) were made as hydrophobic as possible (type C1), the beads on the sides (SC5 and SC2) were made slightly less hydrophobic (type C4), and the beads that partake in base pairing (SC3 and SC4) were made as polar as possible (type P6). All hydrophobic beads also had 'h' labels added to also increase attractive forces between the bases. Once these changes were made (among others), the single strand DNA showed significant base stacking and helical conformation, albeit still relatively unstable compared to AA. The DNA also showed propensity towards base pairing, which was attributed to the large discrepancy of polar and non-polar within the base itself, which is hypothesized to prevent the bases from intercalating. To further improve this model and make the DNA more stable, more research should be done on finding the balance between the backbone strain and the base stacking interactions, with focus on entropic driving forces.

References

[1] S. J. Marrink, A. H. d. Vries and A. E. Mark, “Coarse Grained Model for SemiQuantitative Lipid Simulations,” *The Journal of Physical Chemistry B,* vol. 108, no. 2, pp. 750-760, 2004.

[2] P. C. T. Souza, R. Alessandri, J. Barnoud, S. Thallmair, I. Faustino, F. Grünewald, I. Patmanidis, H. Abdizadeh, B. M. H. Bruininks, T. A. Wassenaar, P. C. Kroon, J. Melcr, V. Nieto and V. Corradi, “Martini 3: a general purpose force field for coarse-grained molecular dynamics,” *Nature Methods,* vol. 18, no. 4, pp. 382-388, 2021.

[3] M. A. Kastenholz, T. U. Schwartz and P. H. Hünenberger, “The Transition between the B and Z Conformations of DNA Investigated by Targeted Molecular Dynamics Simulations with Explicit Solvation,” *Biophysical Journal,* vol. 91, no. 8, pp. 2976-2990, 2006.

[4] G. D. Team, “GROMACS documentation,” [Online]. Available: https://manual.gromacs.org/. [Accessed 11 2023].

[5] G. Bussi, “Canonical sampling through velocity rescaling,” *Journal of Chemical Physics,* vol. 126, no. 1, p. 014101, 2007.

[6] M. Parrinello and A. Rahman, “Polymorphic transitions in single crystals: A new molecular dynamics method,” *Journal of Applied physics,* vol. 52, no. 12, pp. 7182-7190, 1981.

[7] B. Hess, “P-LINCS: A parallel Linear Constraint Solver for Molecular Simulation,” *Journal of Chemical Theory and Computation,* vol. 4, no. 1, pp. 116-122, 2008.

[8] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, “A smooth particle mesh Ewald method,” *Journal of Chemical Physics,* vol. 103, no. 19, pp. 8577-8593, 1995.

[9] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, “ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB,” *Journal of Chemical Theory and Computation,* vol. 11, no. 8, pp. 3696-3713, 2015.

[10] “"OL" Force Field Refinements for RNA and DNA Simulations,” Palacký University Olomouc, [Online]. Available: https://fch.upol.cz/ff_ol/downloads.php. [Accessed 11 2023].

[11] F. Grunewald, R. Alessandri, P. C. Kroon, L. Monticelli, P. C. T. Souza and S. J. Marrink, “Polyply; a python suite for facilitating simulations of macromolecules and nanomaterials,” *Nature Communications,* vol. 13, no. 1, p. 68, 2022.

[12] F. Grunewald, “fgrunewald/Martini_PolyPly,” [Online]. Available: https://github.com/marrink-lab/polyply_1.0. [Accessed 11 2023].

[13] T. Ando and J. Skolnick, “Crowding and hydrodynamic interactions ikely dominate in vivo macromolecular motion,” *Proceedings of the National Academy of Sciences,* vol. 107, no. 43, pp. 18457-18462, 2010.

[14] “Martini 3.0,” 23 8 2022. [Online]. Available: http://cgmartini.nl/index.php/martini-3-0. [Accessed 11 2023].

[15] F. Grunewald, “ fgrunewald/fast_forward,” 9 5 2022. [Online]. Available: https://github.com/fgrunewald/fast_forward. [Accessed 11 2023].

[16] “ChatGPT 4,” OpenAI, [Online]. Available: https://chat.openai.com/chat. [Accessed 11 2023].

[17] A. H. -J. Wang, G. J. Quigly, F. J. Kolpak, J. L. Crawford, J. H. v. Boom, G. v. d. Marel and A. Rich, “Molecular structure of left-handed double helical DNA fragment at atomic resolution,” *nature,* vol. 282, pp. 680-686, 1979.

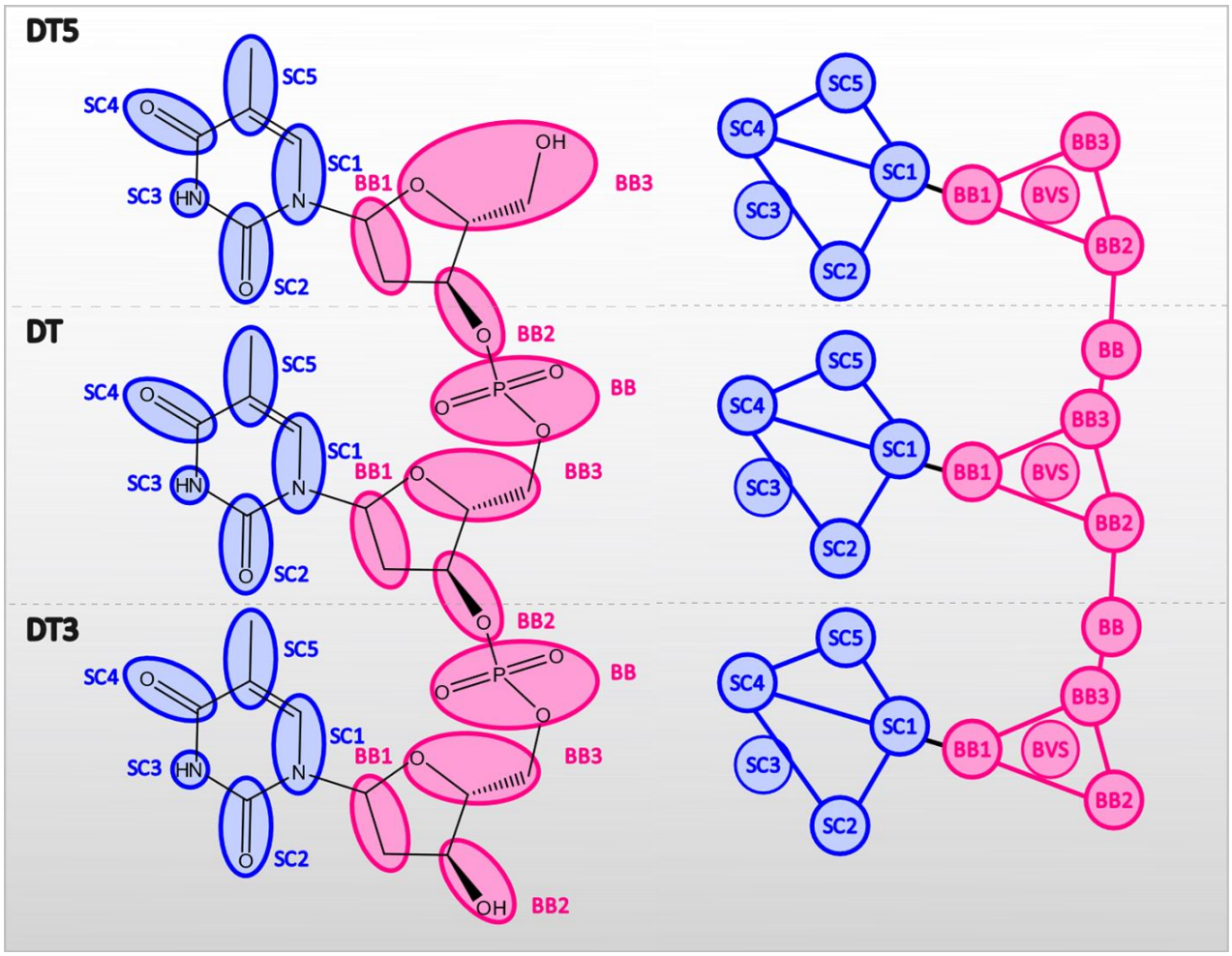
[18] C. H. Mak, “Unraveling Base Stacking Driving Forces in DNA,” *Journal of Physical Chemistry B,* vol. 120, no. 26, pp. 6010-6020, 2016.

[19] A. Dragan, C. Crane-Robinson and P. Privalov, “Thermodynamic basis of the α-helix and DNA duplex,” *European Biophysics Journal,* vol. 50, pp. 787-792, 2021.

[20] V. N. Potaman and R. R. Sinden, “National Library of Medicine,” [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK6545/. [Accessed 22 2023].

Appendix

A 1. Mapping of thymine single strand from all atomistic to coarse grain.



A 2: Tables containing the Bonded Force Field Parameters of adenine and thymine single strands.

Table 3: Bonded Force Field Parameters for adenine and thymine residue bonds and corresponding backbone bonds.

residue	bonds	R_{bond} (nm)	K_{bond} (kJ mol^{-1})	Function type [4]
DA5	BB1_SC1	0.24	8000	1
DA	BB_BB3	0.33	8000	1
	BB1-SC1	0.24	5000	1
DA3	BB_BB3	0.33	8000	1
	BB1_SC1	0.24	8000	1
$Backbone_{DA}$	BB2_+BB	0.305	8000	1
DT5	BB1_SC1	0.24	8000	1
DT	BB_BB3	0.33	8000	
	BB1-SC1	0.24	8000	1
DT3	BB_BB3	0.33	8000	
	BB1_SC1	0.24	8000	1
$Backbone_{DT}$	BB2_+BB	0.26	8000	1

Table 4: Bonded Force Field Parameters for adenine and thymine residue angles and corresponding backbone angles.

residue	angles	θ (deg)	K_{angle} (kJ mol^{-1})	Function type [4]
DA5	BB3-BB1-SC1	100	250	10
	BB2-BB1-SC1	140	250	10
	BB1-SC1-SC2	98	200	10
	BB1-SC1-SC5	160	300	10
DA	BB-BB3-BB2	88	200	10
	BB3-BB1-SC1	100	250	10
	BB2-BB1-SC1	140	250	10
	BB1-SC1-SC2	98	200	10
	BB1-SC1-SC5	160	300	10
DA3	BB-BB3-BB2	88	200	10
	BB3-BB1-SC1	100	250	10
	BB2-BB1-SC1	140	250	10
	BB1-SC1-SC2	98	200	10
	BB1-SC1-SC5	160	300	10
$Backbone_{DA}$	BB3-BB2-+BB	135	250	10
	BB2-+BB-+BB3	105	250	10
DT5	BB3-BB1-SC1	100	250	10
	BB2-BB1-SC1	150	250	10
	BB1-SC1-SC2	85	200	10
	BB1-SC1-SC5	160	300	10
DT	BB-BB3-BB2	103	200	10
	BB3-BB1-SC1	100	250	10

Table 4 (continued):

residue	angles	θ (deg)	K_{angle} (kJ mol^{-1})	Function type [4]
	BB2-BB1-SC1	150	250	10
	BB1-SC1-SC2	85	200	10
	BB1-SC1-SC5	160	300	10
DT3	BB-BB3-BB2	103	200	10
	BB3-BB1-SC1	100	250	10
	BB2-BB1-SC1	150	250	10
	BB1-SC1-SC2	85	200	10
	BB1-SC1-SC5	160	300	10
$Backbone_{DT}$	BB3-BB2-+BB	155	250	10
	BB2-+BB-+BB3	97	250	10

Table 5: Bonded Force Field Parameters for adenine and thymine residue dihedrals and corresponding backbone dihedrals.

residue	dihedrals	Φ (deg)	$K_{dihedral}$ (kJ mol^{-1})	Function type [4]
DA5	SC1_SC2_SC5_SC4	180	400	2
DA	SC1_SC2_SC5_SC4	180	400	2
DA3	SC1_SC2_SC5_SC4	180	400	2
$Backbone_{DA}$	BB_BB3_BB2_+BB	180	100	2
	BB3_BB2_+BB_+BB3	61	50	2
	BB2_+BB_+BB3_+BB2	140	50	9
DT5	SC1_SC2_SC5_SC4	180	400	2
DT	SC1_SC2_SC5_SC4	180	400	2
DT3	SC1_SC2_SC5_SC4	180	400	2
$Backbone_{DT}$	BB_BB3_BB2_+BB	160	100	2
	BB3_BB2_+BB_+BB3	65	50	2
	BB2_+BB_+BB3_+BB2	145	50	9