

DEPARTMENT OF PHYSICS

TFY4235 - NUMERICAL PHYSICS

Exam 2023 - 10039

1 Preliminary

1.1 Terminology

My code was written in Julia.

1.2 Preliminary tasks

- In a structure of 300 monomers there are 299 covalent bonds. If we assume that each bond can take 4 directions, we get 4^{299} different tertiary structures. That is approx. 1.04×10^{180} . Of course there would be a lot of degeneracy in these states.
- It would then take

$$1.04 \times 10^{180} \times 10^{-12} s = 1.04 \times 10^{168} s. \quad (1)$$

Which is some time!

- In a regular computer with integers with 64-bit precision you can store integers up to $2^{63}-1$. I use Julia for my code, and according to their documentation the largest integer value supported is $2^{128}-1$. The highest floating point number is a Float64, which is according to the IEEE 754 format. The largest float is then around 1.8×10^{308} .
- Yes for both. Adding or multiplying with two different types of precision leads to a result with the precision of the least precise type. If the highest

precision number has decimals that the lower precision number does not support, the answer you get could be quite different than from the real answer.

- For single precision one could add 2^{-32} and for double precision the smallest number is 2^{-53} . I tested this in Julia, it is in the bottom of "tests.jl".

2 Tasks

2.1 Task 1

2.1.1 Primary structure

I defined a mutable struct named Acid. Each Acid has a position (a tuple with ints) and a type (an int). For example are (3,0) and (4,0) neighbours. To create a polymer, I made a vector consisting of Acids. Thus the indices in the vector could be used to find covalent bonds between Acids in further tasks. When plotting the structures themselves, the colors are decided by the type of monomer. It is worth noting that the colors are random for each plot. As the interaction energies are randomized at the start of most simulations, I thought it unnecessary to keep the colors constant for different simulations.

2.1.2 Monomer-monomer interaction energies

I did not make a nearest neighbour list for each monomer. Instead, the nearest neighbours were found during simulation when needed. The important factor in making the interaction energy matrix is that it is symmetric, since the interaction energies between monomers are the same no matter the order. For example, the index [15,5] should have the same value as the index [5,15]. An example of an interaction matrix is seen in Figure 1.

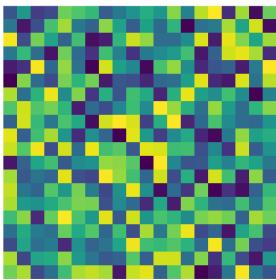


Figure 1: One example of an interaction matrix, plotted as a heat map.

2.1.3 Tertiary structures

Figure 2 shows three different tertiary structures initialized in a folded state. It is observed that the leftmost structure has 6 monomer-monomer interactions, the middle has 4 interactions, while rightmost has 2. The energies for each structure reflect the amount of interactions it has, as it increases in the same order that the number of interactions decreases. In a physical manner, every system is driven to minimize its internal energy. When you then have negative values for nearest neighbour interactions, it implies that there will be a driving force for folding the polymer together and form as many of these bonds as possible, thus having minimal internal energy.

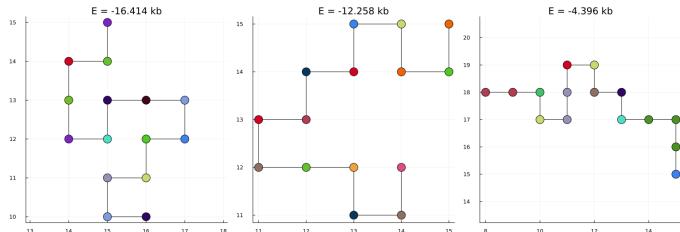


Figure 2: Three different tertiary structures with their energies. The colors indicates the type of the monomer.

2.1.4 Monte Carlo

The logger was not implemented, I instead initialized arrays before the simulation and pushed the parameters in the arrays for each sweep.

2.1.5 Temperature = 10

Figure 3 shows the energy, end-to-end distance and RoG as a function of MC sweeps for a system with 15 monomers with $T = 10$. Simulations with more than 500 steps were also conducted, however the same chaotic behaviour of the parameters was observed for steps up to 10 000. The tertiary structures are also shown in the figure, at 300, 400, and 500 MC sweeps. It is observed that the tertiary structures have notable changes over time.

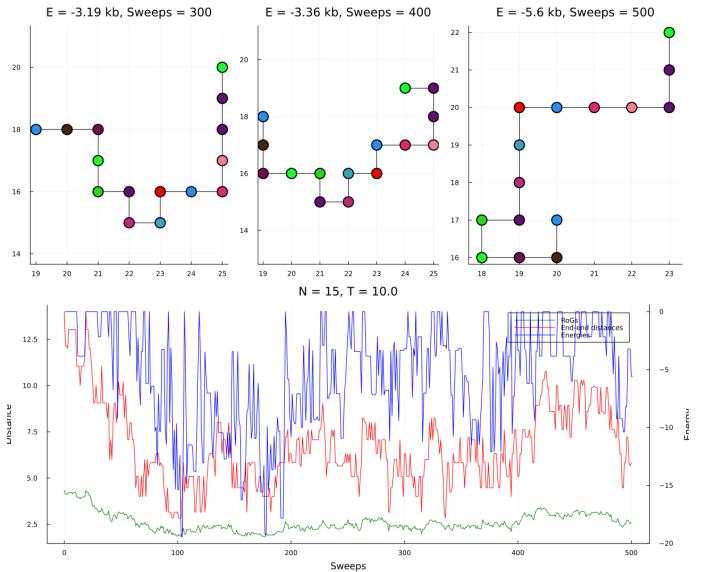


Figure 3: Top: Polymer structure consisting of 15 monomers during simulation with $T = 10$.
Bottom: The energy, end-to-end distance and RoG as a function of MC sweeps for the polymer.

2.1.6 Temperature = 1

Simulations with $T = 1$ were first done for 100 and 1000 MC sweeps, where some of the structures seemed to converge already in the first tens of steps. After simulating for 10 000 sweeps, however, it was observed that a majority of the structures found themselves in meta-stable states in the beginning of the simulations. Figure 4 shows two systems, one simulated for 500 sweeps and the other for 10 000 sweeps. The first seems to find a steady state after around 200 sweeps, while the other found several meta-stable states before it minimized itself after around 5000 sweeps. The time to find one steady state is thus only in the tens of sweeps, but if one wants to consider the global minima it can take several thousand sweeps. From a physical point of view, this is a result of the low temperature. As temperature lowers, transitions that do not result in a lower internal energy in the system are less probable. So if the system finds itself in a meta-stable state with an energy barrier, it would take a longer time (on average) with small temperatures to overcome the energy barrier (break the bonds) and try new structures

that might result in lower internal energy. The tertiary structures in Figure 4 are thus observed to have smaller variations in time than in Figure 3

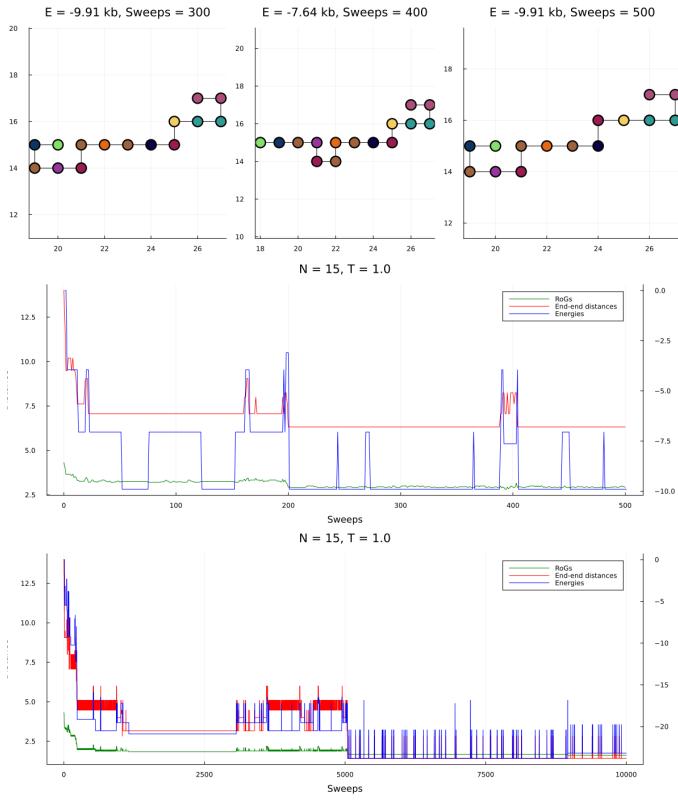


Figure 4: Top: Polymer structure consisting of 15 monomers during simulation with $T = 1$. Middle: The energy, end-to-end distance and RoG as a function of MC sweeps for the same polymer as the top. Bottom: Same plot as middle, but for a new primary structure with same parameters and more MC sweeps.

2.1.7 Phase diagram

- (a) The trend was observed to be that it takes a longer time to find equilibrium as T increases. For $N = 15$ the plots for T larger than 2 are chaotic similar to Figure 3, which is logical as the small number of monomers makes the tertiary structures too simple to "hold together" when the T increases. Defining an equilibrium state for these plots is hence difficult. For $N = 50$ and 100, however, it is seen that energies up to 4 T converge at a steady state, see Figure 5 and Figure 6. These polymers have monomers to create more bonds, which makes them more stable at higher T . At small temperatures and small N , steady states are found almost immediately, as seen before for $N = 15$. $N = 50$ also found a meta-stable state almost immediately. For $N = 100$ in Figure 6, however, it does not find a steady state until 1000 sweeps. It finds meta-stable states, however much shorter than for $N = 50$ and 15. This is because the

size results in far more combinations of structure, so even if one state is stable there is a high probability to transition to another state with lower energy, since the theoretical number of stable states is so high.

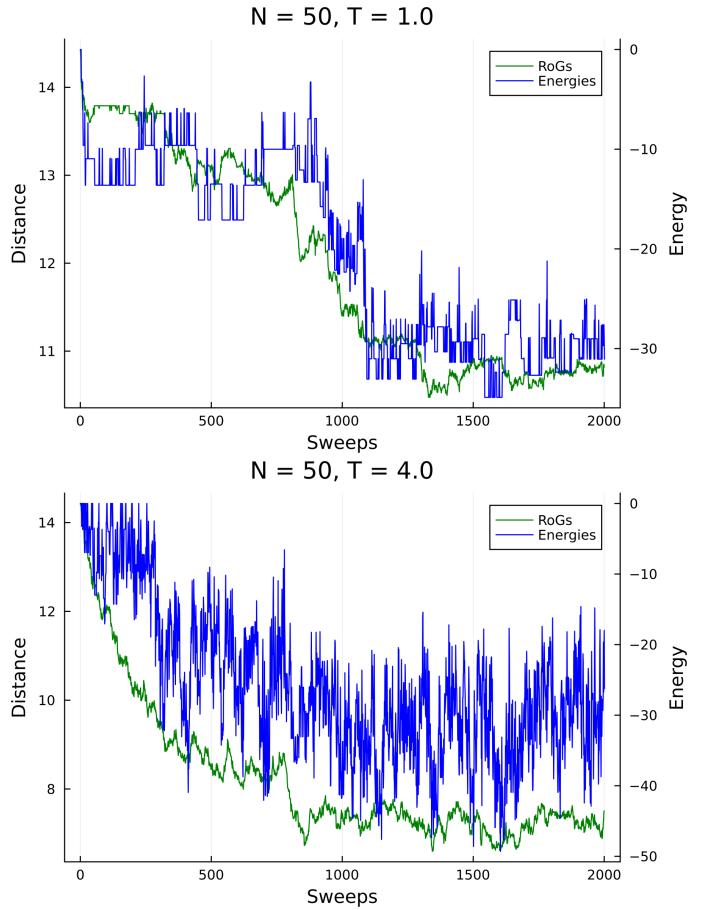


Figure 5: Energy and RoG of 50 monomers for 2000 MC sweeps at different temperatures.

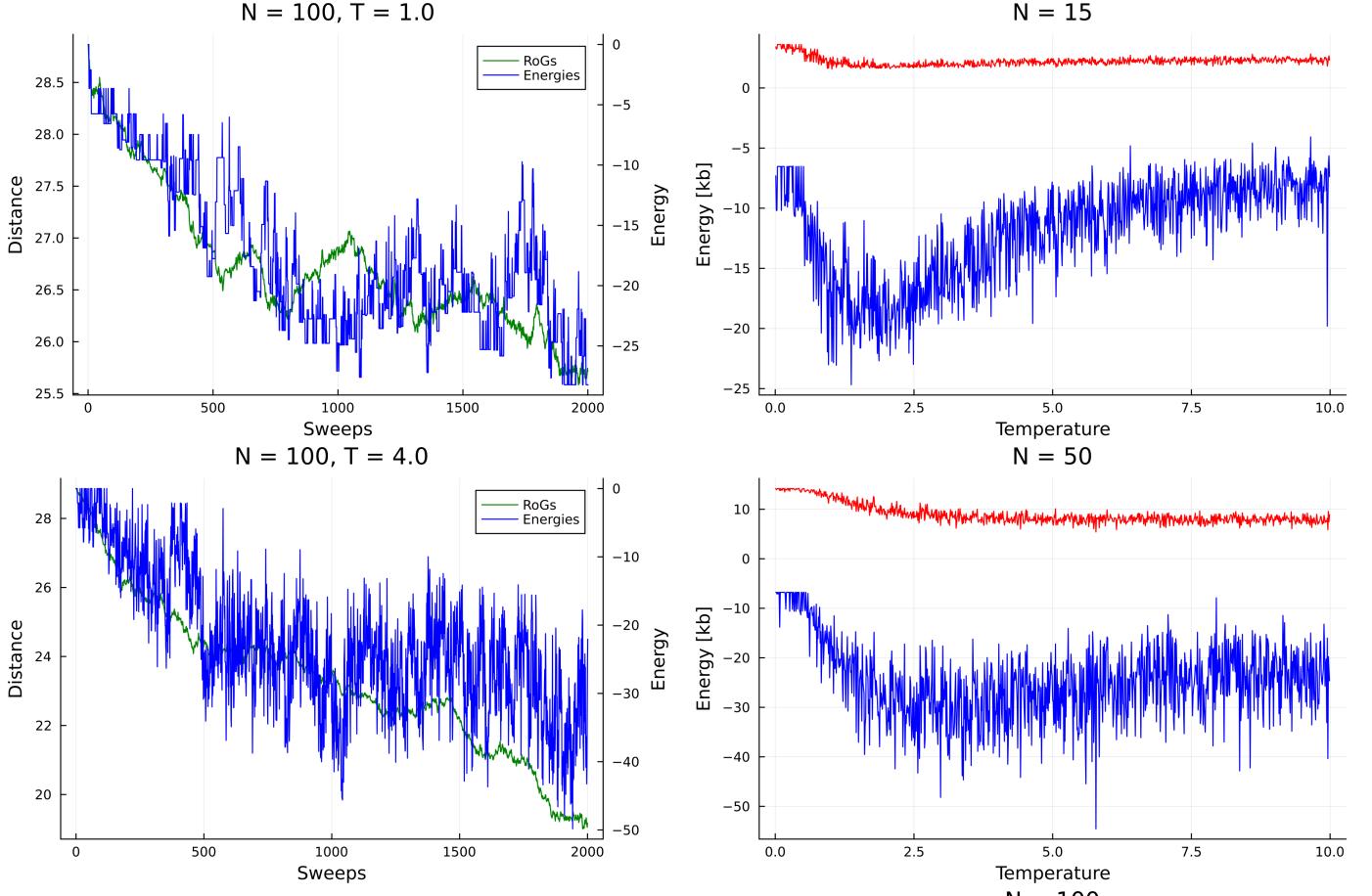


Figure 6: Energy and RoG of 100 monomers for 2000 MC sweeps at different temperatures.

- (b) Simulations for $N = 15$ and 50 were conducted with 1000 temperatures ranging from 0.01 to 10 . Each simulation lasted for 2000 MC sweeps. The simulation took far too long for $N = 100$, so this was done with 100 temperatures ranging from 0.1 to 10 for 2000 sweeps. The results are seen in Figure 7. There was a bug in the Plots package used for plotting and the data structures that made it impossible to make two separate y-axis for distance and energy in this case, so the y-axis show both energy and distance for the energies and RoGs respectively. This also lead to the plot not being with decreasing T .

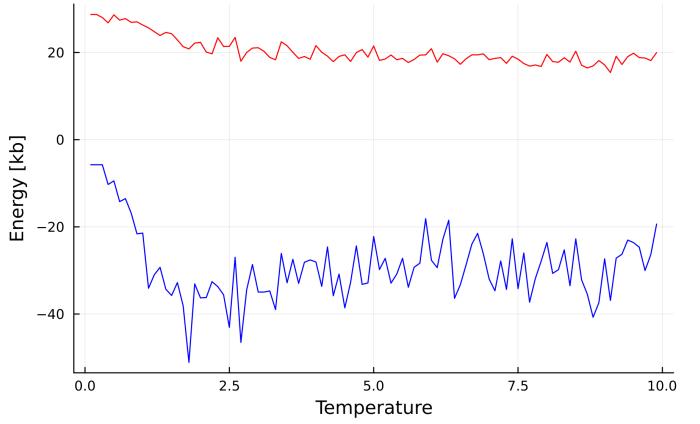


Figure 7: Phase diagrams $E(T)$ (blue) and $RoG(T)$ (red) for different N . Y-axis has values of energy and distance.

The starting point was chosen by simulating several systems and finding the average point where the systems usually found steady states. Since for most systems it required quite many sweeps to find this, it was set to 1000 sweeps, as this was where most plots found their steady state. The examples Figure 5 and Figure 6 also show this behaviour. Another method for doing this would be to divide the plot into intervals, and calculate the mean of these intervals. Then a starting point would be where the difference in mean for two succeeding intervals was

lower than a given value. Thus it could be chosen individually for each plot. However, since the plots had large fluctuations for increasing T , and a trend of flattening out after 1000 sweeps, making an educated guess was believed to not affect the results too much. If there was more time, a sophisticated solution could be implemented to get even more accurate results.

- (c) The overall structure of the phase diagrams are similar; the energies and RoGs decrease slowly with temperature until the critical temperature where they increase rapidly. The critical temperature seems to be a bit lower for $N = 15$ than for the other two. For $N = 15$ it is around 1.5, while for $N = 50$ and 100 it lies closer to 2. The plot for $N = 15$ has less fluctuations, a logical result as the combinations of structures possible is much lower than for larger N . This makes it easier to pin point the critical temperature than for the other two. It could have been useful to simulate with more MC sweeps to decrease the fluctuations for larger N .

2.1.8 Simulated annealing

- (a) Figure 8 shows two different structures simulated for 10 000 sweeps. Their energies were calculated as the mean of the last 500 sweeps, and the fluctuations as the standard deviation in the same interval. The energies are quite different, but the energy fluctuations are almost identical. This is reasonable, as there are several randomized factors in the simulation, making it harder for some structures to find the lowest energy state. But for 10 000 sweeps with $T = 1$ it is likely the structures ended up in very stable states. This could explain why the deviations are so similar. Even though the interaction matrix is randomized at the start of each simulation, there are 2000 different bonds, the values are in a small interval, and they are chosen from a uniform distribution. Thus each structure should on average have very similar bond strengths. And since the temperature and number of sweeps is the same, it would be of high probability that both structures found themselves in states were the energy barriers were around the same.

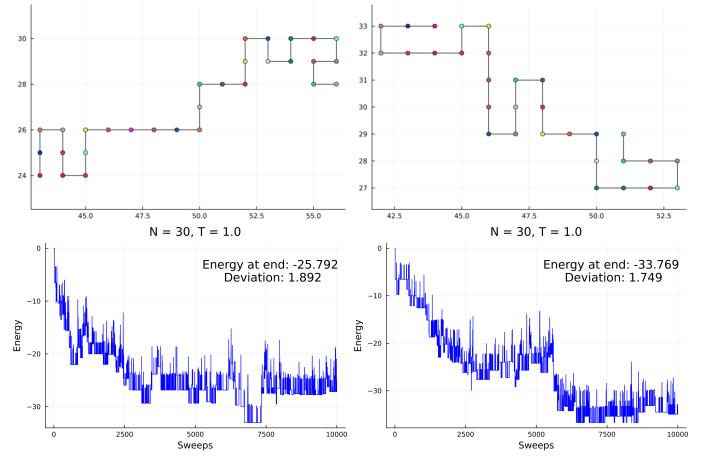


Figure 8: The structures, energy, and energy deviation of two simulations done with $T = 1$

- (b) As seen in Figure 9, the structure found with simulated annealing (SA) seems to be much more stable than the other two in Figure 8. Almost every single monomer has at least one nearest neighbour interaction, leading to a much lower energy of $-44 k_b$. It points to a success in finding the global energy minimum. The deviation of zero is a result of the temperature range, it started at 4 and ended at 0. The temperature was too low for transitions to happen after it reached the final structure. What is interesting is that it took over 6000 sweeps for the structure to stabilize, which is comparable to the result in Figure 8. It would be interesting to test different starting temperatures and see if it is faster. One would assume it was faster to anneal, however the results only show that the ending state had lower energy.

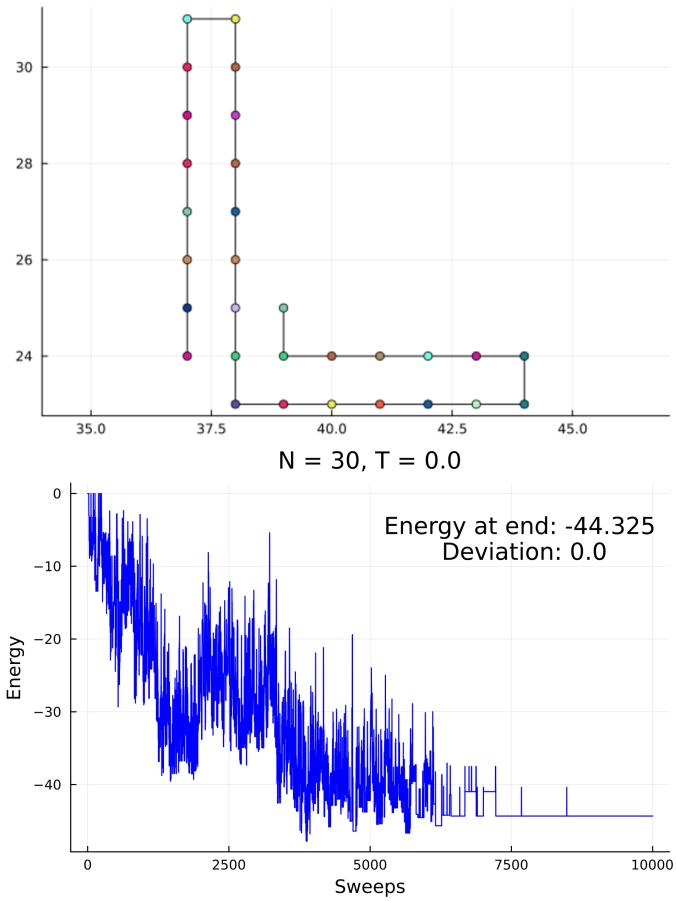


Figure 9: The structure, energy (in k_b , and energy fluctuation for one simulated annealing process. Temperature started at 4 and ended at 0.

(c) Levinthal's paradox tells that even though the time to test each conformation of a protein would be longer than the lifetime of the universe, regular proteins find their correct conformation in times down to microseconds. The complexity of the energy landscape can be seen in all results in this exam. There are almost no processes where the energies have a positive or negative gradient for a long amount of time. Meta-stable states slow down the process immensely, one can for example see in Figure 4 that the system was in a meta-stable state for 5000 sweeps until it found a lower energy state. The proteins have therefore found the correct "pathways" through the energy landscape to arrive at the states that are needed. The process is complicated, and the solution to the paradox is that proteins use guides to direct them within "funnel-like energy landscapes" [1].

2.1.9 Positive interactions

Figure 10 shows the structure, energy and energy fluctuation of a 10 000 sweep simulation with 50 monomers.

The interaction matrix is also seen as a heat map in the figure. The new structures are not as compact as the ones seen in Figure 9. It is observed that none of the pairs are nearest neighbours, and the polymer itself is relatively unfolded. The ends are almost exactly the same shape, and it seems like they have tried to fold together in the most compact form that does not involve positive interaction bonds.

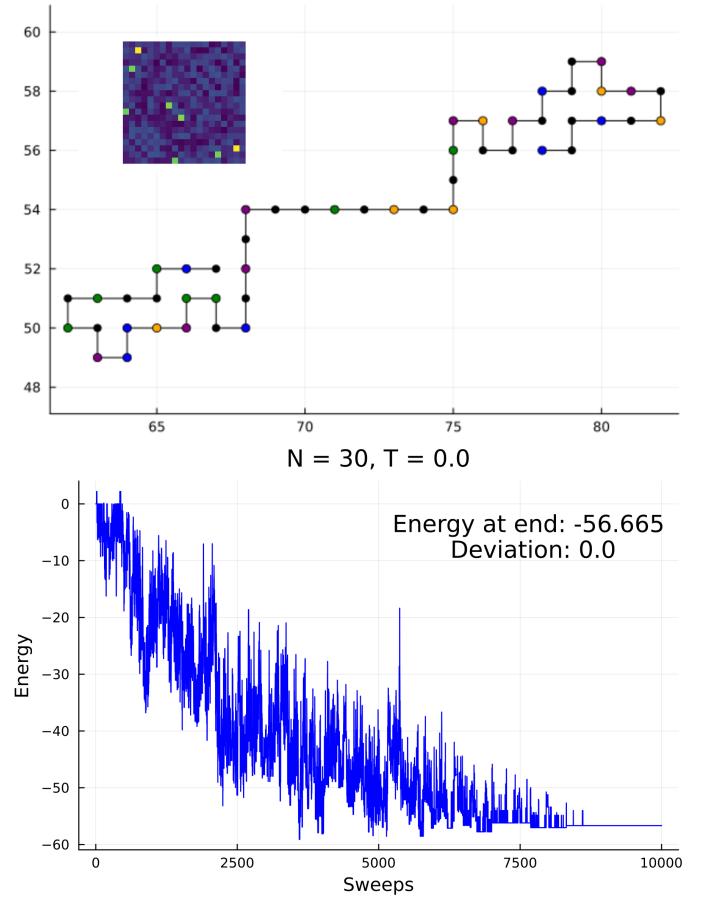


Figure 10: Simulation of 50 monomers where 5 pairs in the interaction matrix have positive sign. The pairs that have positive interactions with each others have the same colors in the plot. The interaction matrix is shown at the top as a heat map.

2.2 Task 2

2.2.1 Extend to 3D

There was not much difference in extending the system. When checking for nearest neighbours the functions were updated to check the dimensions of the position tuples. Some functions were copied and specified for 3D, such as plotting and finding the radius of gyration. An amino acid at the end of the structure can jump to 4 lattice points, instead of the 2 when in 2 dimensions. An internal acid however, can still only jump to one spot. This

is because of the restriction, they can only jump in two dimensions each time, or else they would break the rigid bonds.

2.2.2 Evolution of 3D protein

Figure 11 shows a simulation of a 3D protein with $N = 15$ and $T = 10$.

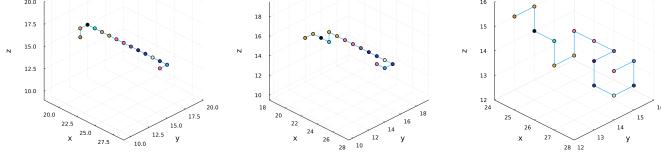


Figure 11: Simulation of a 3D protein with $N = 15$ with $T = 10$. From the left, $x = 1, 10, 100$ sweeps.

2.2.3 Phase diagrams in 3D

- (a) Figure 12 and Figure 13 show 3D simulations with the same parameters as Figure 5 and Figure 6. Especially for $N = 50$ the results show that it took a longer time to find equilibrium in 3 dimensions. This was a trend for all values of N and T , but there are exceptions. For example, it only took 500 sweeps $T = 1$ in Figure 13 to reach a steady state, but this can be explained by the randomness of the simulations. In general, it took more time for the 3 dimensional simulations to find equilibrium. The cause for this is believed to be the larger variation of structures possible in 3D, that the polymers will on average try more structures before it reaches a steady state, simply because there are more structures to try.

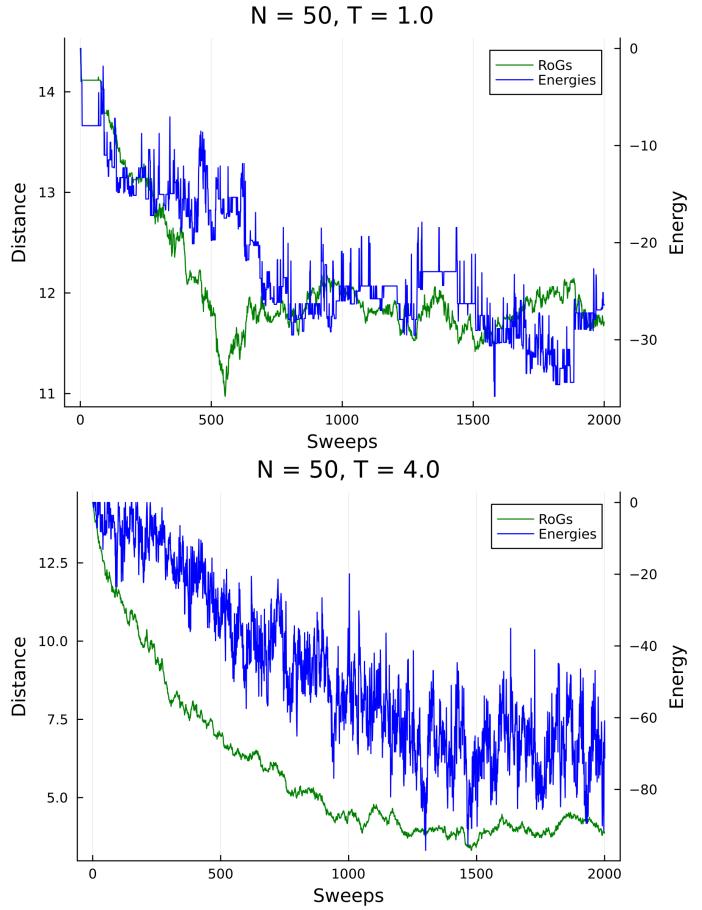


Figure 12: Energy and RoG of 50 monomers in 3D for 2000 MC sweeps at different temperatures.

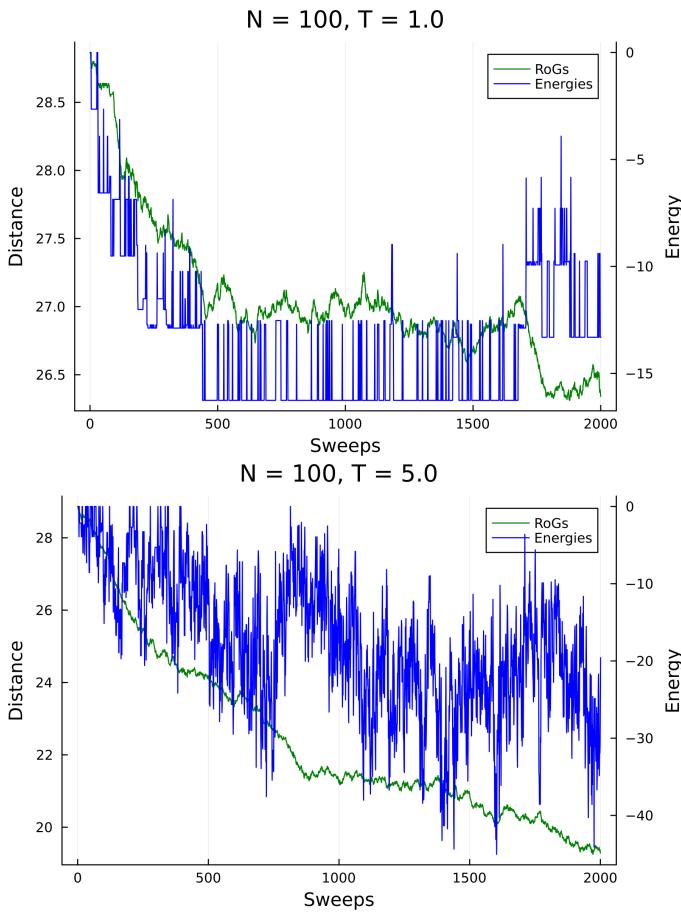


Figure 13: Energy and RoG of 100 monomers in 3D for 2000 MC sweeps at different temperatures.

for taking the averages was increased to 1200 sweeps instead of 1000 sweeps.

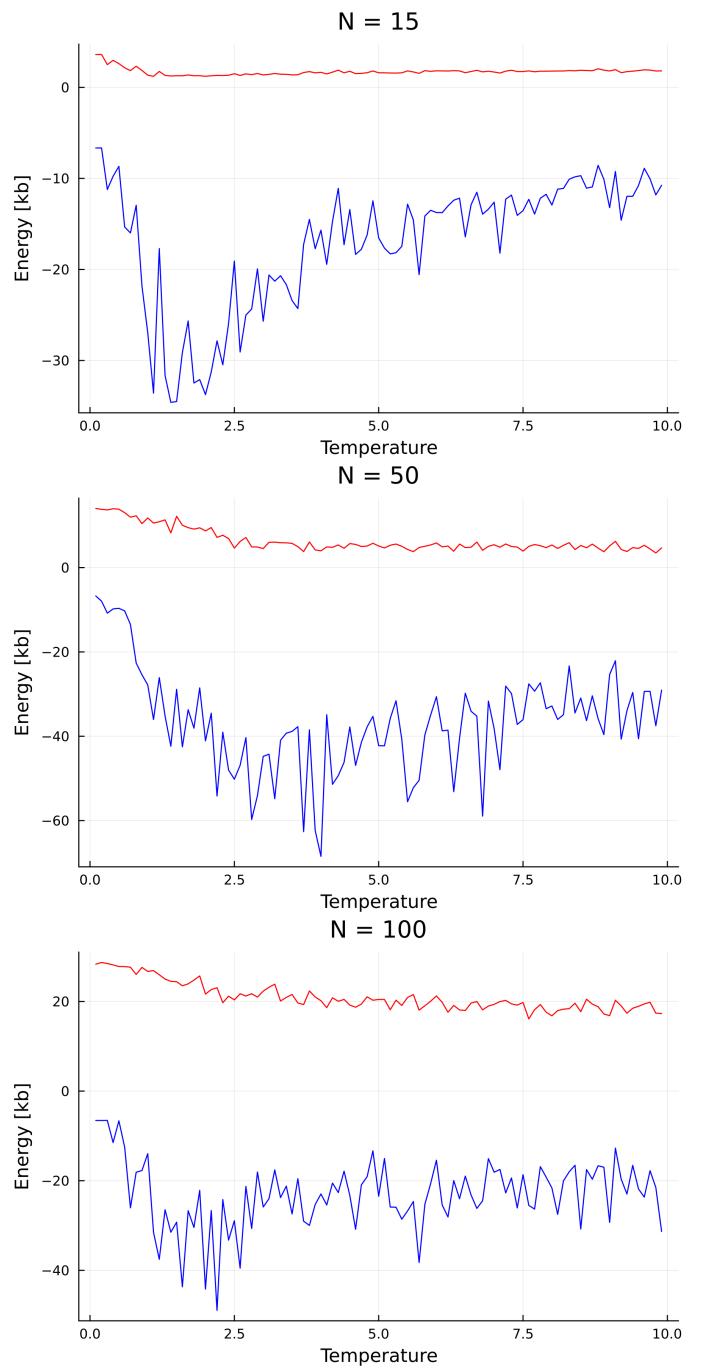


Figure 14: Phase diagrams $E(T)$ (blue) and $\text{RoG}(T)$ (red) for different N in 3D. Y-axis has values of energy and distance.

Bibliography

- [1] Ken A Dill and Hue Sun Chan. ‘From Levinthal to pathways to funnels’. In: *Nature structural biology* 4.1 (1997), pp. 10–19.

Appendix

I worked heavily with the candidates 10076 and 10069. They did not write their code in Julia, but we discussed the overall structure and methods to find the solutions.

The code for this exam can be messy in some areas. I found that Julia had some performance problems concerning vectorization, meaning that it was faster to explicitly write out for example every element in a tuple in an operation. This turned the code into a whole mess, and resulted in the simulations giving wrong answers. So I turned back to the slow method, but there might be some artifacts left of the optimization that I have not found. In addition, the main function has several lines commented out, these are only lines that were used when testing and plotting different data during development of the code. The tasks.jl file is also quite messy, but none of the actual logic is implemented here, only different ways to simulate and plot the results.