Norwegian University of Science and Technology

# Project 2, TMA4320: Protein Folding

Stu.nr: 478679 Stu.nr: 478694 Stu.nr: 478704

Feb/Mar 2018

## 1 Introduction

Protein folding is a biophysical process where the chains of amino acids in the protein-molecule (also called monomers) keeps spontaneously folding itself til the molecule reaches a state with minimized potential energy (relative to the point of reference of maximal 0 J; an unfolded state). A molecular state in this context is defined as one of the possible shapes while folding.

In this document, we study a simplified problem. The protein is assumed to live in two-dimensional space, and is only allowed to make right angled ($90^o$) folds, or *twists*, around a single monomer. This yields a grid-like structure. It is considered illegal for the protein to intersect itself, i.e. for two monomers to occupy the same place on the grid.

It is desirable to not let the numerical algorithm skew the results in one direction or another. That is, to not let the chosen protein twists be determined by a controllable algorithm. Therefore, all protein twists are determined by randomly generated numbers (a so-called *Monte-Carlo method*). A twist performed to a protein is calculated from the previous state, so the numerical algorithm is in fact a *Monte-Carlo Markov process*. Given that the random number generator is sufficiently random (which we assume it is), this should yield the most likely state for the protein after sufficiently many twists.

The commentary in the sections below aims to answer the questions posed in the project description. Figures along with the commentary represent numerical simulations, and the simulations are constrained to the simplifications made. Still, the simplified simulation can tell us a great deal about the actual physical situation.

## 2 Answers to questions

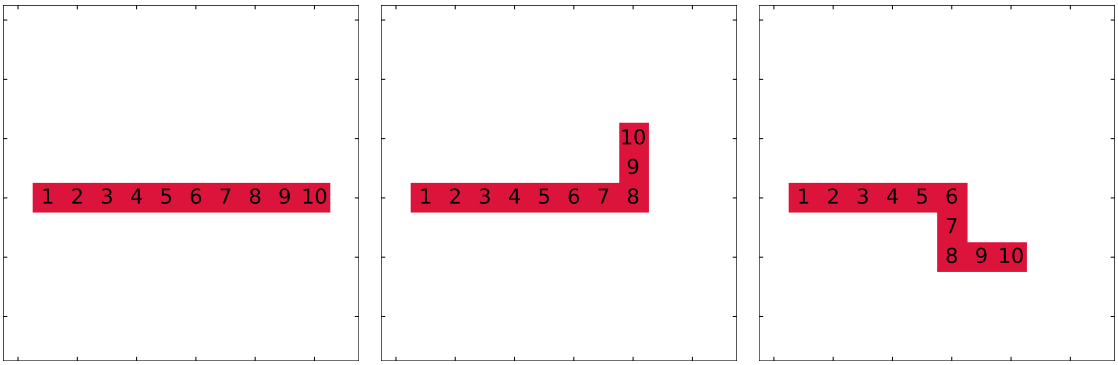
### Question 1

#### 1.1

To represent a protein of length $N$, we chose to initialize a square "empty grid" (a zero matrix) of size $N+2$, and fill it with ascending integers corresponding to where adjacent protein monomers are located (i.e. the first monomer in the chain is represented by the number 1, the second by 2 etc.). By choosing the midpoint of the protein to be a fixed point, no possible rotation of the protein can exceed the grid.

#### 1.2

To check if a twist is legal (it is legal if the resulting protein does not intersect itself), we extract all the monomers that are to be twisted into their own matrix, rotate it, and check if the places they would end up are empty. If the twist is legal, the execution is similar. In figure 1, we show that the initialization and rotation of the protein works correctly.




No twists One twist Two twists

Figure 1: Exercise 1.2. Starting with a straight polymer consisting of 10 monomers, two twists are performed. First, a counter-clockwise twist around monomer number 8. Second, a clockwise twist around monomer number 6.

### Question 2

#### 2.1

We are interested in the mean energy of the protein at different temperatures. For temperatures in the range between 0 and 5000 K, a new protein consisting of 15 monomers

is created. It is then twisted a sufficient amount of times for the protein to reach a fairly stable state.

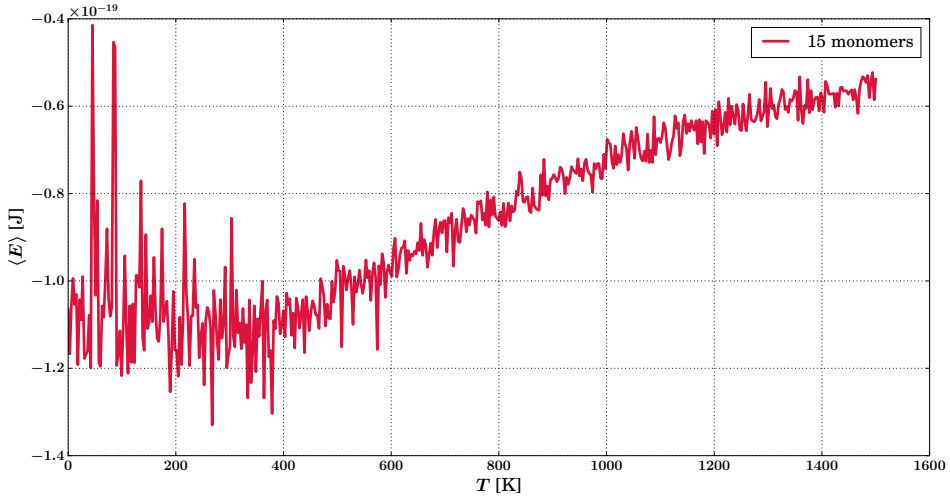



Figure 2: Question 2.1. The mean energy, $\langle E \rangle$, of a 15-monomer long protein, twisted sufficiently many times for each temperature.

The mean energy as a function of temperature can be seen in figure 2.

*Sufficiently* many twist to ensure a fairly stable state, $d(T)$, was chosen by using the formula

$$d(T) = d_{\max} e^{-sT}, \tag{1}$$

where

$d_{\max}$: the number of necessary twists at $T = 0$ K
$s$: decides the rate of change of the number of twists necessary.

This formula reflects the fact that for low temperatures, and thus fewer thermal fluctuations, more twists are used to reduce the extent to which proteins get caught in local energy-minimums for the binding energy. For the plot in figure 2, $d_{\max} = 13000$ and $s = 5 \cdot 10^{-4}$ K$^{-1}$ was used. For $T = 0$ K, this yields 13000 twists. This should be sufficient to let the protein reach minimal energy in most (but not all) cases for low temperatures. Then, for $T = 1500$ K, just over 6000 twists are performed. Because more thermal fluctuations are present for higher temperatures, it is unlikely that a protein gets stuck in the same state for too long. Thus, performing more twists does not make it more likely for the mean energy to approach any particular value, and there is no point

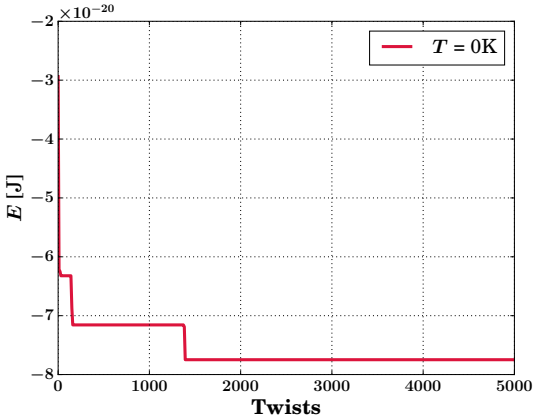



Figure 3: Question 2.2. The binding energy of a single protein twisted at $T$=0 K as a function of the number of twists.

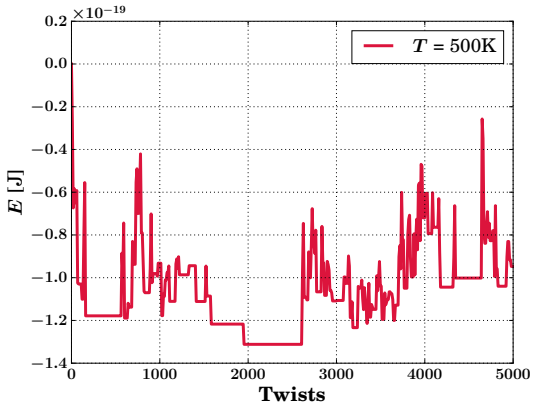



Figure 4: Question 2.2. The binding energy of a single protein twisted at $T$=500 K as a function of the number of twists.

in performing too many twists. 6000 should be more than enough to reflect the average value $\langle E \rangle$ takes on somewhat accurately.

## 2.2

We now want to consider if the formula (1) for $d$ is justified.

The binding energy for two different proteins are plotted in figures 3 and 4 for two different temperatures $T$.

One observes from the figures that the protein that twists at $T = 0$ K reaches a minimum fairly quickly, and performs no jumps in energy. The protein that twists at $T = 500$ K has a lot more apparent randomness connected with its binding energy. The reason so many twists are performed at low temperatures, is to give the protein enough time to find its minimum value. Even for $T = 500$ K, the binding energy does not seem to approach any value. This is our justification to perform fewer twists at high temperatures, and the formula for $d$ gives the amount of twists consistently with this.

## 2.3

Considering figure 2, one observes that $\langle E \rangle$ increases steadily for temperatures $T \gtrsim 200$ K. The reason for this could be thermal fluctuations giving rise to more spontaneous transmissions to higher energy states. The simulated ratio between probabilities of different states equals the Boltzmann factor $\exp(-\beta \Delta E)$ for $\beta = 1/k_B T$ and $\Delta E =$

$E_i - E_j$, where $E_i$ and $E_j$ is the energy in state $i$ and $j$ respectively. Thus, the Boltzmann factor increases when $T$ increases.

When $T$ approaches 1200-1300 K, it appears that the slope $d\langle E\rangle/dT$ decreases. This is expected to happen at some point, as the mean energy could never exceed the maximum possible energy state (the unfolded state).

### 2.4

For low values of $T$, we observe from the plot in figure 2 that $\langle E\rangle$ varies wildly for small changes in $T$. As demonstrated in figure 3, the protein will rarely jump to a higher energy state at low temperatures. Thus, the occasional high mean energy is likely a case of a protein getting caught in a local minimum point for the binding energy. For each new temperature, a new protein was defined, so the protein had a new chance to avoid the local minimum in the next plotting point.

### 2.5

We repeat the procedure from section 2.1 on a protein consisting of 30 monomers. The plot of the mean energy, $\langle E\rangle$, is given in figure 5.

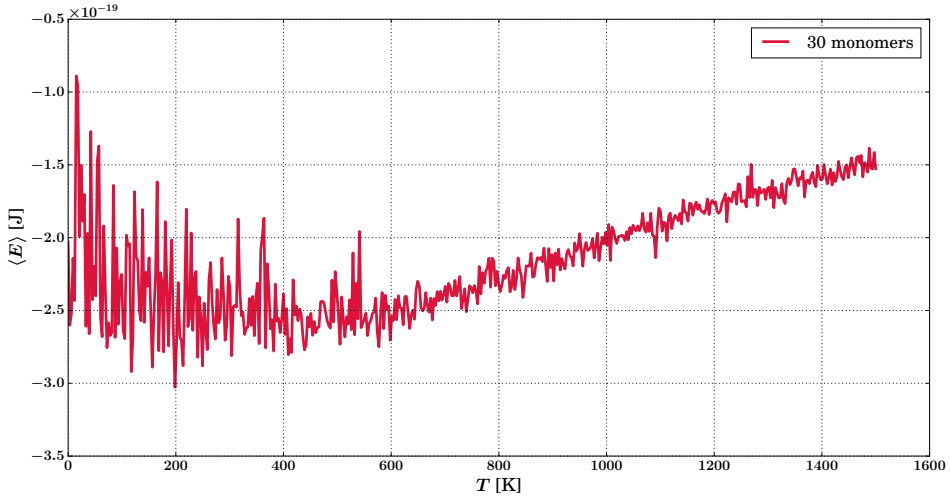



Figure 5: Question 2.5. The mean energy, $\langle E\rangle$, of a 30-monomer long protein, twisted sufficiently many times for each temperature.

As the protein is now twice as long as before, the plot is made with $d_{\max} = 26000$ in equation 1, which is twice as high as before. This yields well over 20000 twists for low

temperatures. The other parameter, $s$, was chosen to be 0.001, as this yields around 6000 twists for $T = 1500$ K. As discussed previously in exercise 2.1, this should be more than enough, even for a longer protein. By choosing the value of $s$ to be higher for 30-monomer proteins than for 15-monomer ones, we also reduce the run time for the calculations by a not insignificant amount.

The fact that the protein is twice as long, makes the rotation algorithm around four times as complex. This is because each twist makes a rotation to a matrix that represents part of the protein. If the protein length doubles, the expected length of the twisted part of the protein also doubles, and the square rotation matrix then has to have four times the area. We also perform more twists on the longer protein, as it has more chances to get caught in local energy-minimums when rotating at low temperatures. In addition, each illegal twist (which never gets attempted by the algorithm) also needs to execute a rotation of a four times larger matrix, and a longer protein gives more chances that a twist is illegal.

Summarizing; a protein with double length increased the run time for plotting $\langle E \rangle$ by about six times. The larger rotation matrix and increased amount of rotations makes this reasonable, as we are not quite sure about the complexity of the rotation algorithm.

## Question 3

The mean diameter of the protein at different temperatures follows a similar pattern as that of the mean energy. The mean diameter $\langle L \rangle$ is plotted against $T$ for both a protein consisting of 15 and 30 monomers in figure 6. The protein is twisted $d(T)$ times, given by equation 1.

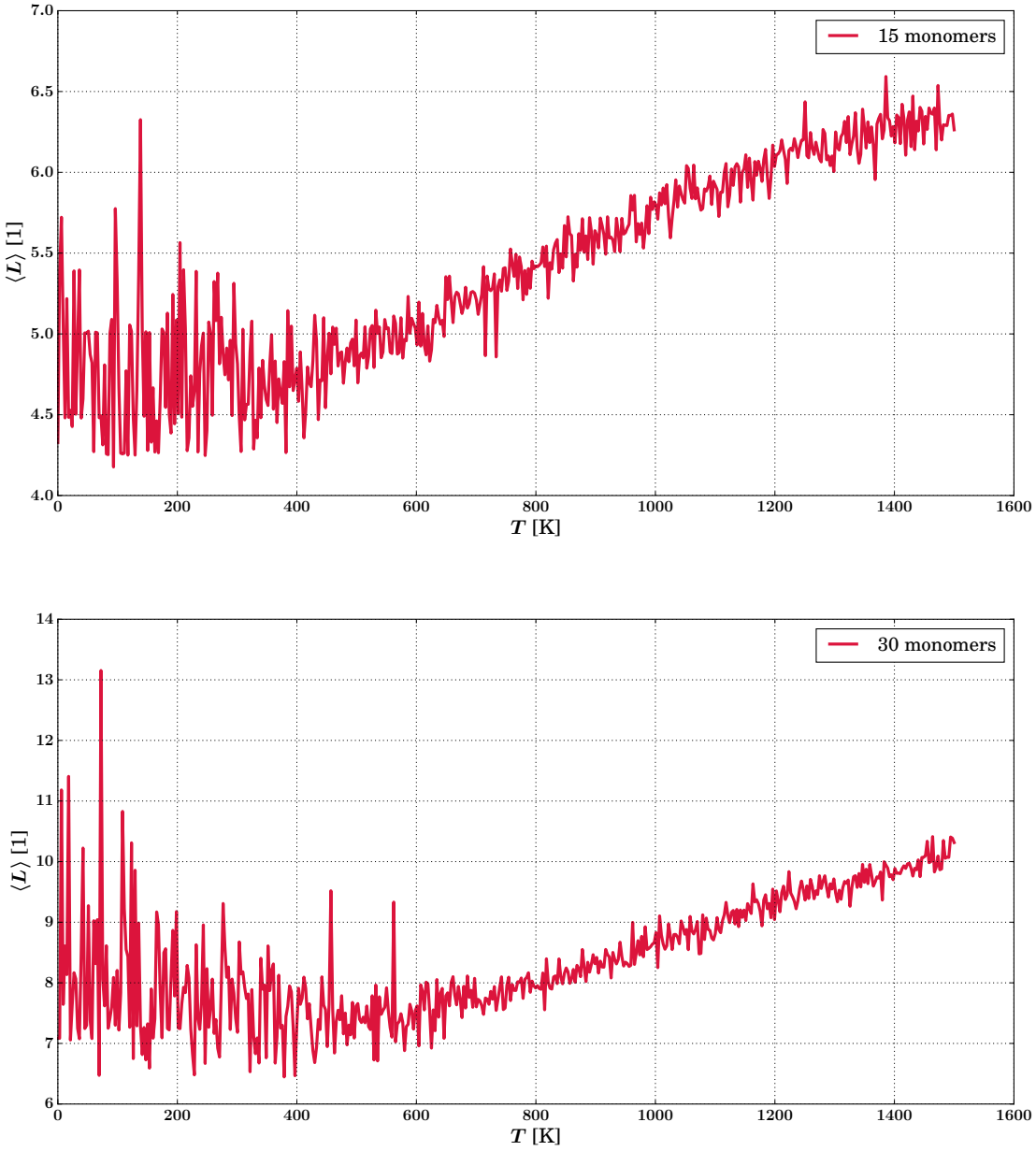



Figure 6: Question 3. The mean diameters of two proteins consisting of 15 and 30 monomers, respectively, as functions of temperature. Each monomer is defined to have length 1.

The diameter of a protein is defined to be the longest distance between any pair of monomers. In the plot, each monomer is defined to have length 1.

The behaviour in the plot is somewhat expected, as the energy of the simulated protein only decreases because of weak bindings between adjacent, non-sequential monomers. The only way to increase the protein's energy is to break some of the weak bindings, which often correlates to increasing the diameter. If the protein has a large diameter, it is likely that only very few weak bindings are present. The smallest possible state

for the protein is an approximate disk shape, and this shape requires numerous weak bindings. For lower temperatures, the lack of thermal fluctuations can result in the polymer randomly being twisted into a local minimum of energy with a large diameter compared to the smallest possible diameter. This is the case in the graphs in figure 6, where there are several spikes of larger mean diameters for temperatures under 200 K.

## Question 4

In this exercise we are interested in the behaviour of a protein during the following cooldown process: Starting at 1500 K, the temperature is decreased by 30 K every 600 random twists, ending at 0 K. Unlike previous exercises, we do not reset the protein to its initial position for each new temperature level. By examining how the mean energy $\langle E \rangle$ and the mean diameter $\langle L \rangle$ changes during the cooldown, one can analyze how the stability of the protein changes with respect to the temperature. In this exercise we will specifically examine proteins with 15 monomers and 30 monomers.

### 4.1

Figure 7 at the top shows the binding energy, $E$, during the cooldown process of a protein with 15 monomers.

At high temperatures and low amounts of twists, the binding energy seems quite random. However, after about 10000 twists, where the temperature is around 1000 K, it seems to start declining, both in terms of the value of $E$ and the apparent randomness between each twist. Around 20000 twists, the temperature is about 500 K, and there is again a noticeable decrease in both the value of $E$ and the variance between twists. This pattern continues all the way to the end of the process. Like one should expect, energy ceases to jump to higher levels when the temperature is very low. Near or at $T = 0$ K, the protein has settled in to a stable state.

The physical interpretation of this behaviour is that when the temperature decreases, the protein is less and less likely to escape a local energy-minimum point it was caught in. For each time the temperature decreases past a point where the protein can no longer (with reasonable probability) escape a local minimum, it seems plausible that this local minimum leads to the eventual global minimum.

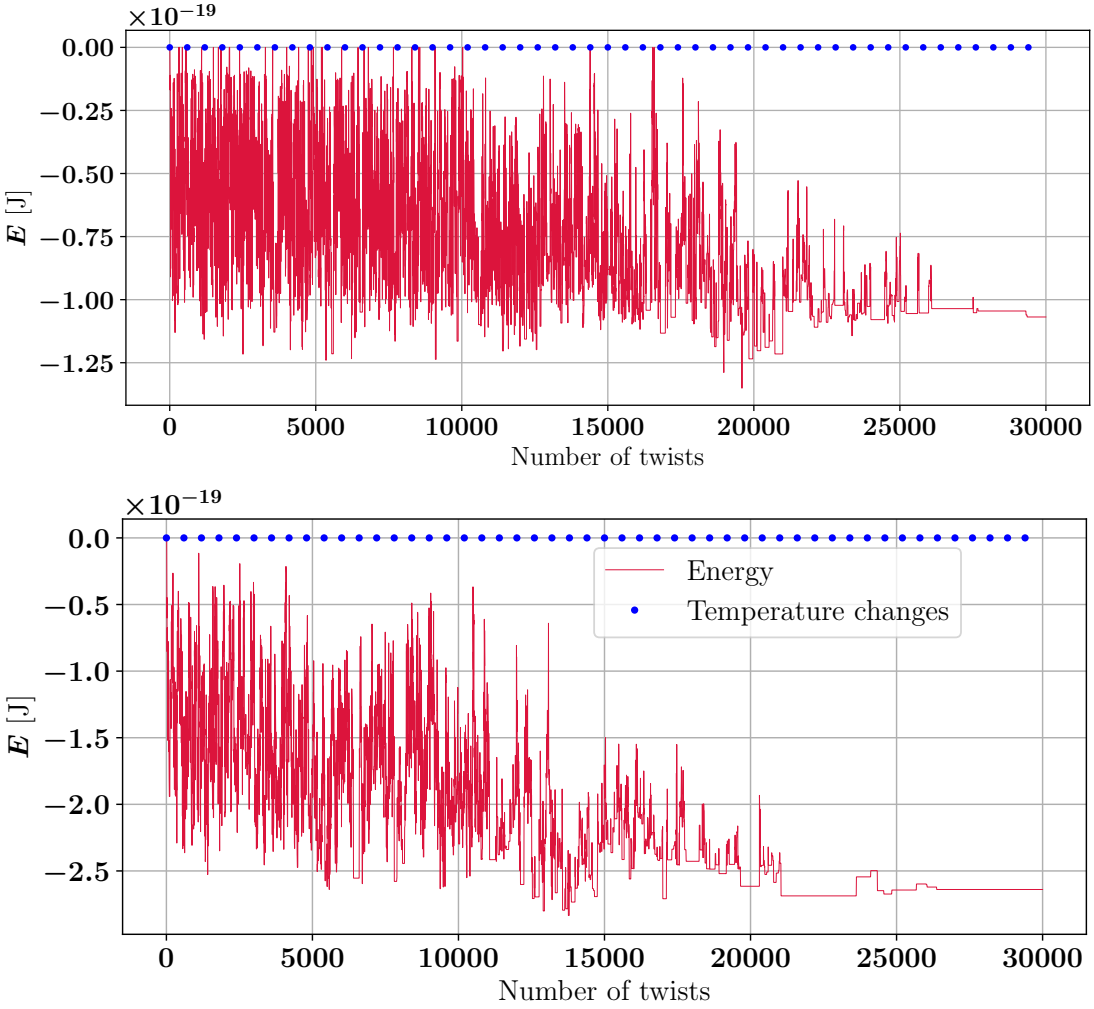



Figure 7: Question 4.1. The plot at the top shows the binding energy $E$ of a polymer with 15 monomers as a function of the number of twists performed, and the plot at the bottom shows an equivalent plot of a protein with 30 monomers. Along with the twisting there is a gradual decrease in temperature from 1500 K to 0 K. The blue dots locates where in the process the temperature changes.

One should also notice that the energy reached in the cooldown-process (from figure 7) about $-1.0 \cdot 10^{-19}$ J, which is lower than the energy reached for performing twists at constant temperature $T = 0$ K (from figure 3), but is higher than the minimum reached by performing twists at $T = 500$ K (from figure 4). Thus, in this case, the cooldown-process did not yield the global minimum, but a local one.

Figure 7 at the bottom shows the same cooldown-process for a protein of length 30. The process is quite similar, but notably, the energy level needs fewer twists (only around 21000 as compared to around 26000) to reach a quite stable state. Again, it does not appear like the protein settled in the global minimal energy state in this particular case, but we do not deny that such a cooldown *could* yield the global minimum.

The fact that this protein has more monomers, also makes it less likely that the ends

of the monomer chain would be in the perimeter of the protein. Thus, there are likely fewer legal low-energy jumps remaining for this protein when it reaches around 20000 twists, and the thermal fluctuations might not be enough to bridge the gap. This would explain the main difference between the plot for the 15-monomer and 30-monomer length proteins.

### 4.2

We now examine how the mean energy, $\langle E \rangle$, in the polymer changes with temperature. As specified in the introduction of exercise **4**, the temperature interval and the amount of twists is $1500 - 0$ K and 600 respectively.

Figure 8 shows the mean energy $\langle E \rangle$ of a protein with 15 monomers. As the temperature decreases, the protein's energy is reduced by the twisting, and the probability that the protein will achieve a high energy state with thermal fluctuations decreases. Near $T = 0$ K, the graph shows that the protein is stuck in a local energy-minimum. As the protein is highly twisted by this point, it will also be highly unlikely to perform new twists that further reduces its state energy. This is also consistent with figure 8.

Compared to $\langle E \rangle$ during a heat up for a protein with 15 monomers (figure 2), the graph of the mean energy near $T = 0$ K behaves quite differently. This is expected because the protein starts out in a maximal energy state (unfolded), and the protein will after some twists enter a state with lower energy.

Similarly, figure 9 shows the mean energy $\langle E \rangle$ of a protein with 30 monomers. As before, the plot suggests that the longer protein reaches a minimum for higher temperatures (i.e. fewer twists). The relatively close relation between $E$ and $\langle E \rangle$ for low amounts of twists validates the explanation in the previous segment for explaining the differences between the plots for the different protein lengths.

### 4.3

The mean protein diameter $\langle L \rangle$ during the cooldown process from 1500 K to 0 K is plotted as a function of temperature in figure 10.

The plot shows a gradual decrease in the mean diameter until around 300 K, where it ends up with a mean diameter of around 4.25 monomer-lengths. As in question 3, one could expect the plot for $\langle L \rangle$ to be quite similar to $\langle E \rangle$, and this is the case (as in that the rightmost 1/6th of the graph in both cases has settled down). The only way to change the binding energy, often involves changing the diameter.

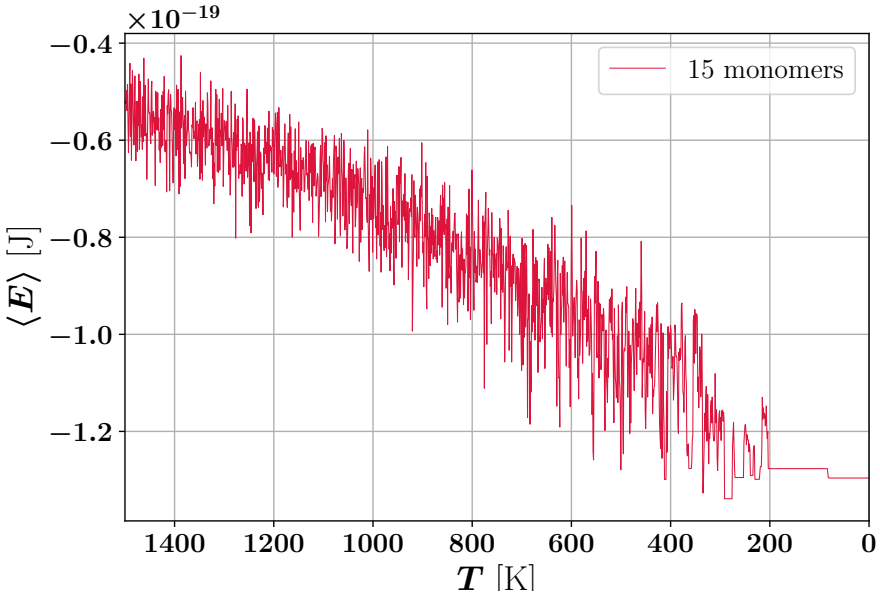



Figure 8: Question 4.2. The mean energy, $\langle E \rangle$, as a function of temperature of a polymer with 15 monomers, following the cooldown process.

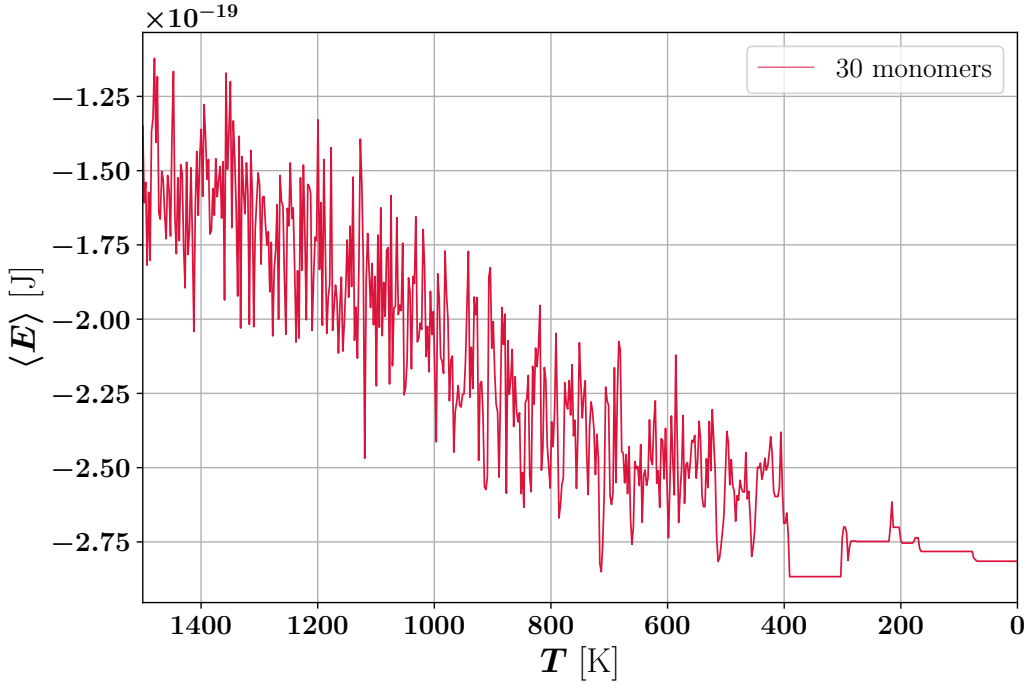



Figure 9: Question 4.5. The mean energy, $\langle E \rangle$, as a function of temperature of a polymer with 30 monomers, following the cooldown process.

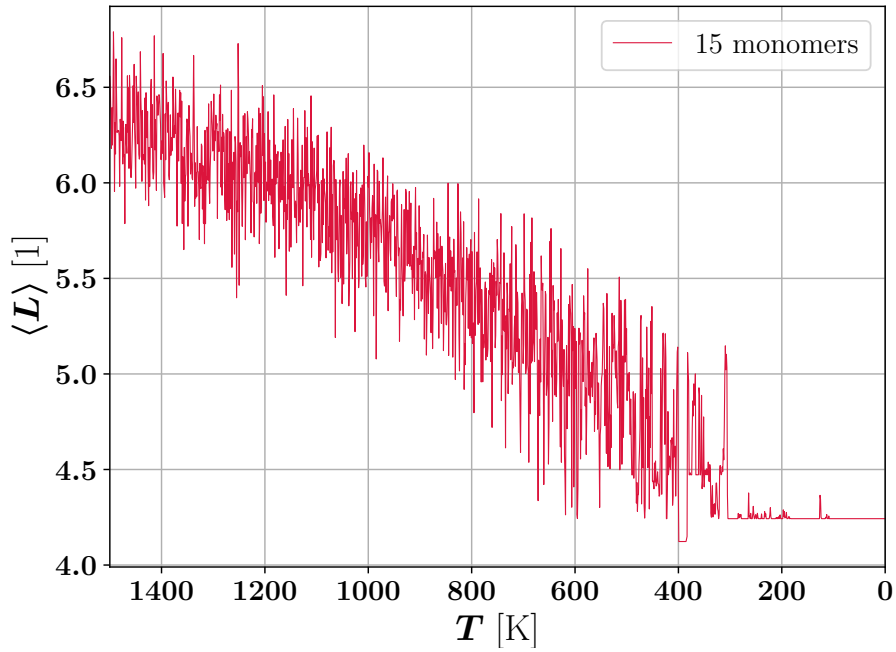



Figure 10: Question 4.3. The mean diameter $\langle L \rangle$ of a protein with 15 monomers as a function of temperature for the cooldown process.

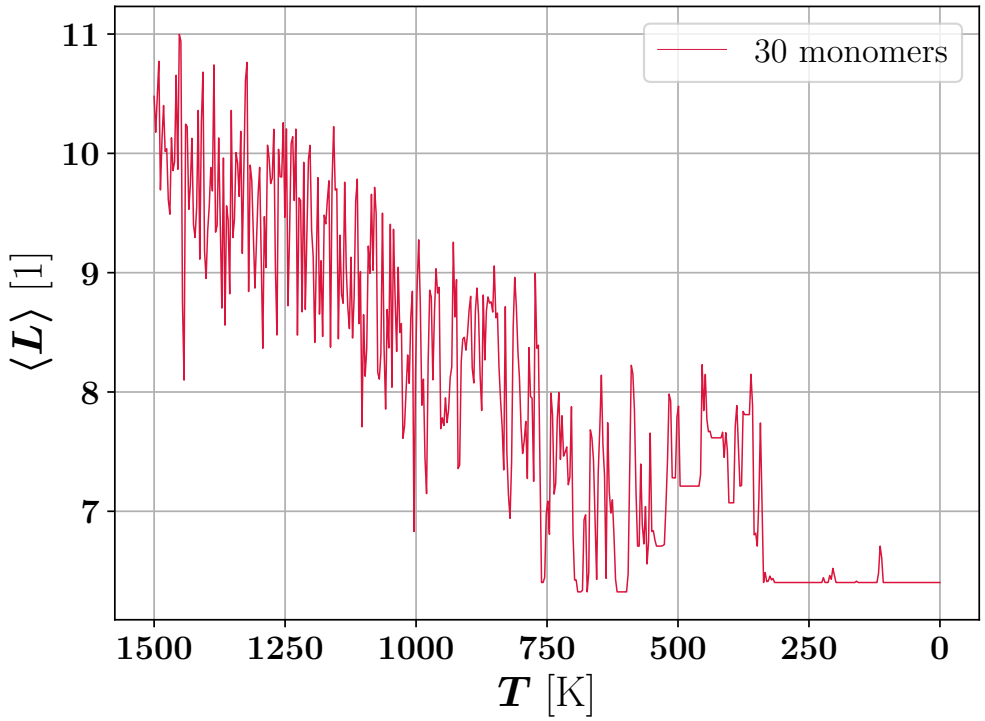



Figure 11: Question 4.5. The mean diameter $\langle L \rangle$ of a protein with 30 monomers as a function of temperature for the cooldown process.

The mean diameter for a protein of length 30 is plotted in figure 11. Unexpectedly,

there is no clear difference when the 15-monomer and 30-monomer protein settles to a stable value for $\langle L \rangle$, as we would expect the graphs of $\langle L \rangle$ to look very similar to the graphs of $\langle E \rangle$ (exercise 4.2) which differ between 15 and 30 monomers. However, the difference is not very significant and could simply be a consequence of the random nature of the simulation.

### 4.4

After performing the cooldown-process a few times, we chose the 30-monomer protein shown in figure 12 to illustrate the end state.

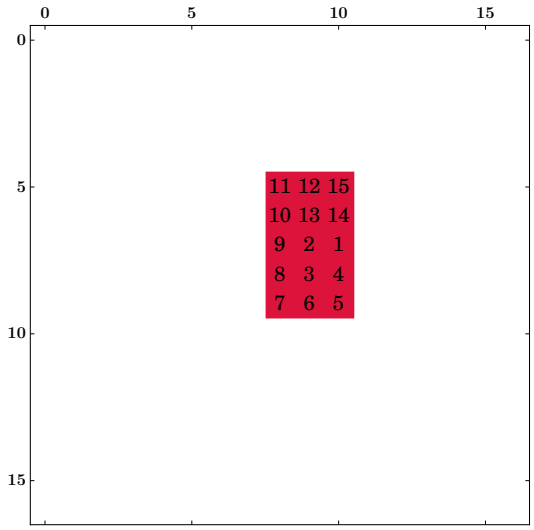



Figure 12: Question 4.4. A visualization of a polymer with 15 monomers after the cooldown-process, at its final state ($T = 0$ K).

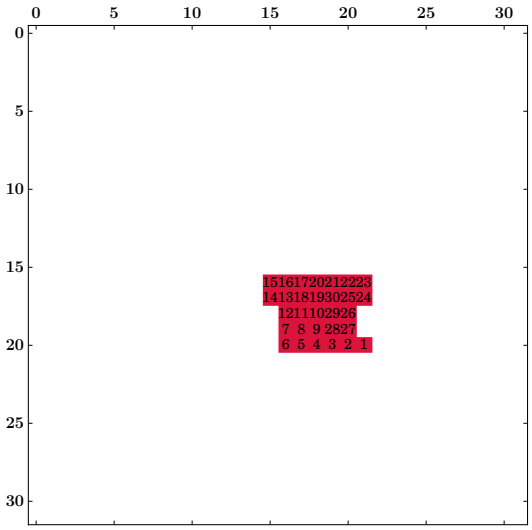



Figure 13: Question 4.5. A visualization of a polymer with 30 monomers after the cooldown-process, at its final state ($T$=0 K).

This protein is twisted to a rectangle-shape, where most monomers have at least one non-sequential adjacent monomer. The protein likely got trapped in this state, because any legal twist would increase the binding energy. Very low values of $T$ would not allow the protein to transition over the energy gap, while slightly larger $T$ would for instance allow a clockwise twist around monomer number 14 in some cases. In other words, the protein is limited to smaller and smaller twists, and this looks to have been the case for the creation of the protein shown in figure 12.

The final state for a 30-monomer protein after the cooldown process is shown in figure 13. As expected, it is fairly *compact*, but it is clear that it is not the globally minimal

energy state.

### 4.5

The discussion for proteins of length 30 have been mixed in with the other answers to question 4.

## 3 Closing words

We used a Monte-Carlo Markov process to simulate protein folding on a two-dimensional grid. The process assumed that the protein folding naturally tends towards low-energy states, but random thermal fluctuations could allow transitions to higher energy states.

The results suggests that a slow cooldown of a protein would generally allow it to enter a lower-energy state than a series of twists at a low temperature all together (as seen in question 2.1, these experiments on proteins would often let them enter local energy minimums).

For example, eggs contain a lot of proteins. A raw egg is liquid, and we can interpret this as the proteins being stuck in (or close to) the global minimum energy state. When heating the egg, we allow the proteins to "straighten out" by thermal fluctuations, and in the cooldown process the proteins gets stuck in lower local energy minimums, hence the egg hardens.