Monte Carlo Simulations for Modeling Protein Unfolding (using end and corner moves)

Link for the screen recording of a protein unfolding:
https://drive.google.com/file/d/14wIIJcwR1eIWHLKQMUFIAmLseKR7OCME/view?usp=sharing

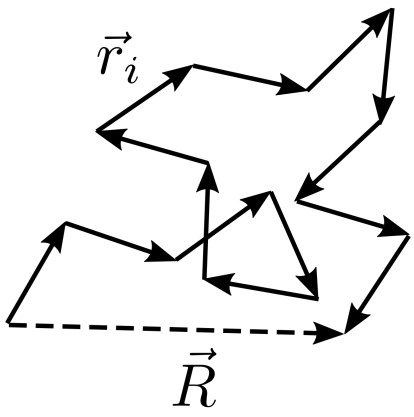
Three different reaction parameters can be used for monitoring the unfolding runs of the protein model, namely, radius of gyration, total interaction energy and end to end distance. These can be used to varying degrees of efficacy due the type of data and the no. of accessible data points each of these variables can take.

Radius of gyration

It is defined by:

$$< R_G^2 > \; = \; \frac{1}{N} \sum_{i=1}^{N} < (R_i - R_{CM})^2 >$$

The radius of gyration of the given protein shows good variability over the course of a single run and it can take values in a continuous spectrum allowing minimal loss of information.



End-to-End distance

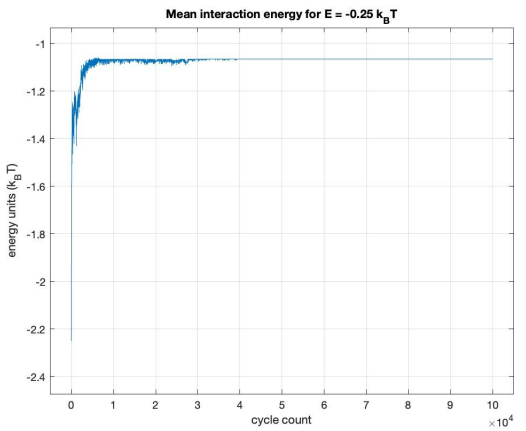
It is defined as the distance between the 1st and the 16th residue for this model of a 16-mer.
This can be used as a crude definition/representation of the size of the molecule.
Although, as we can see in the figure alongside, it is not a very accurate representation of the size of the polymer for highly folded polymers. The folding can cause the ends to come very close together and misrepresent the size of the polymer. Hence we take this parameter with a pinch of salt.

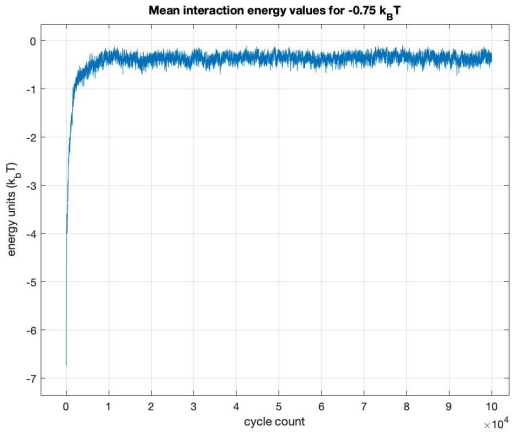
Total Interaction energy:

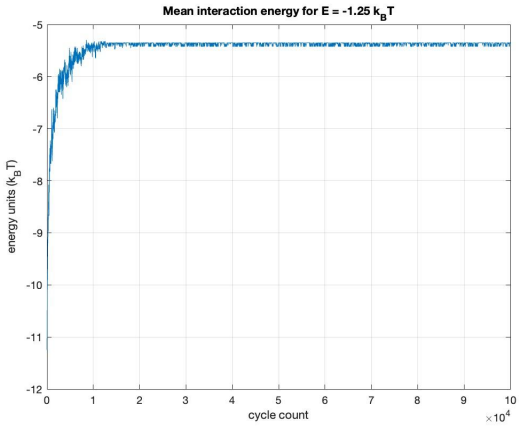
For the simulation runs, the total interaction energy for the polymer has been calculated for all the non-covalent bonds available for the residues in the initial folded state only. Note that two residues which are not close to each other in the initial state of the polymer do not have any interaction even when they come close to each other over the period of the simulation.

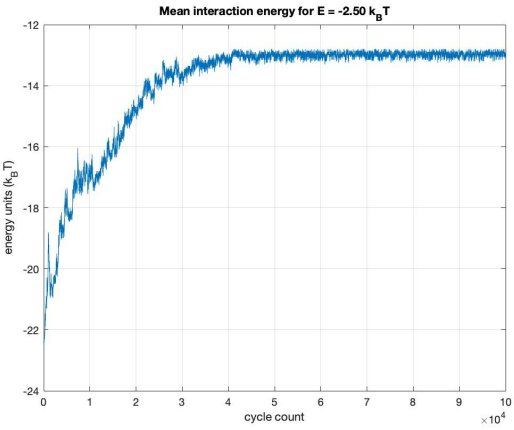
Following are the analyses of the data obtained from the monte carlo simulations run for $50 * 10^5$ cycles.

Mean interaction energy plots for each cycle in the 50 repetitions:









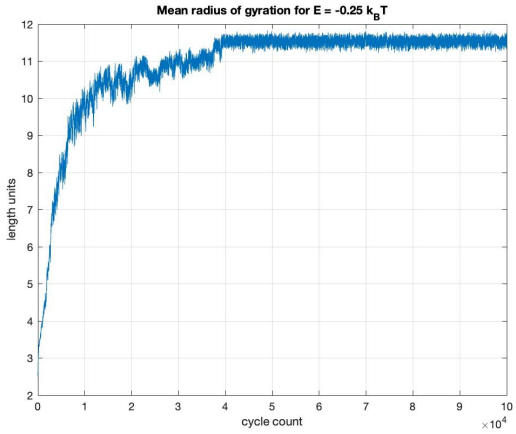
*The scale factor on the y-axis is 0.1 units

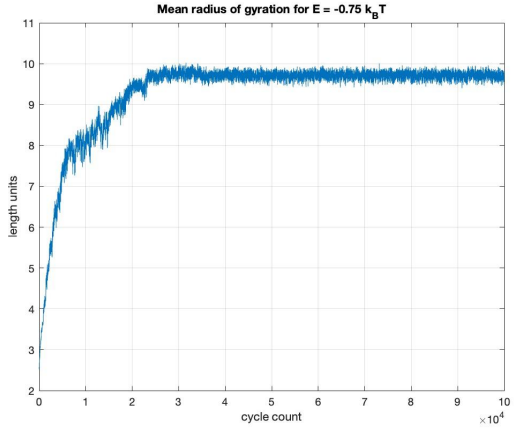
Analysis of the interaction energy data:

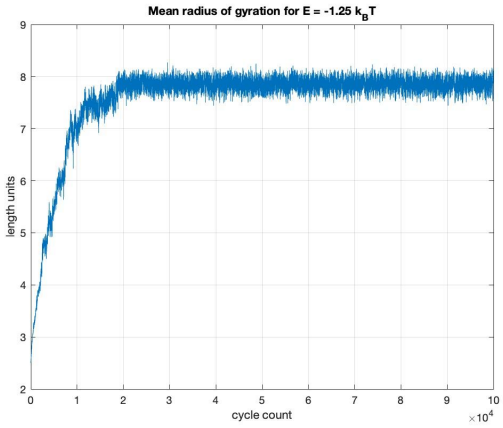
Each data point in the graphs shown are, as mentioned above, the means of each energy state of corresponding cycles in each of the 50 repetitions of the Monte Carlo simulation runs.

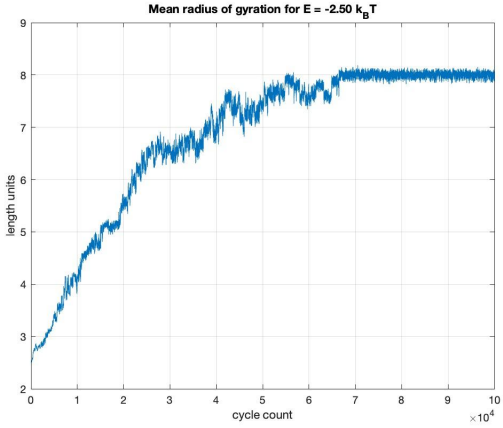
The lower the interaction energy per non-covalent bond of the molecule, the slower it unfolded. The number of cycles on the average over 50 repetitions to reach 0 kT ± (interaction energy per bond) increases. The protein doesn't completely unfold in the -1.25 and -2.50 come to an unfolded energy state lower than 0.

Mean gyration radius plots for each cycle in the 50 repetitions:







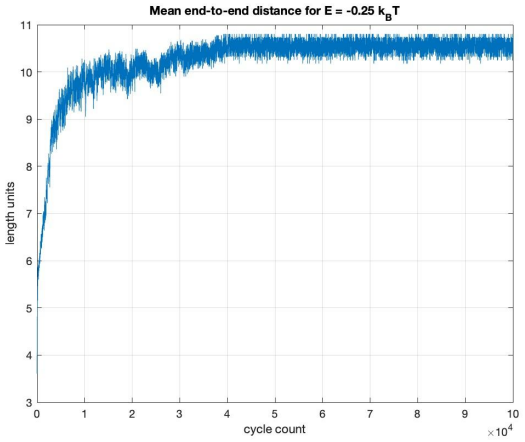


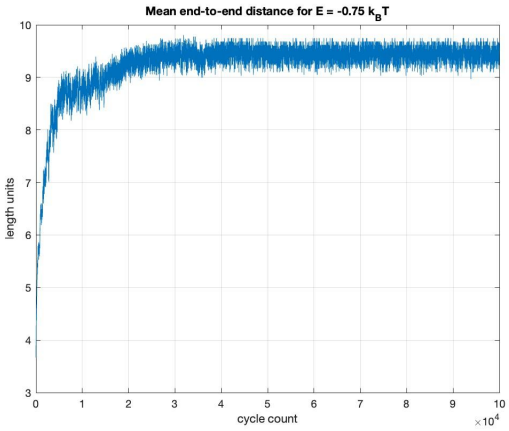
Analysis of the radius of gyration of the protein model:

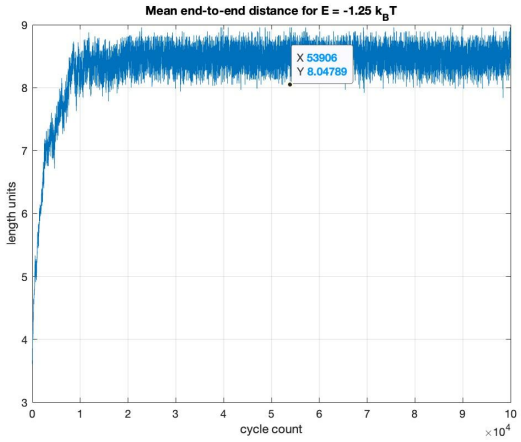
Radius of gyration serves as a better parameter to characterise protein folding due to its ability to take a larger spectrum of values in the same range for each energy value. This is also a true measure of the size of the molecule as it gives information of the relative positions of each residue in every state. The radius of gyration values have also been given the same averaging treatment as in the previous dataset for interaction energies.

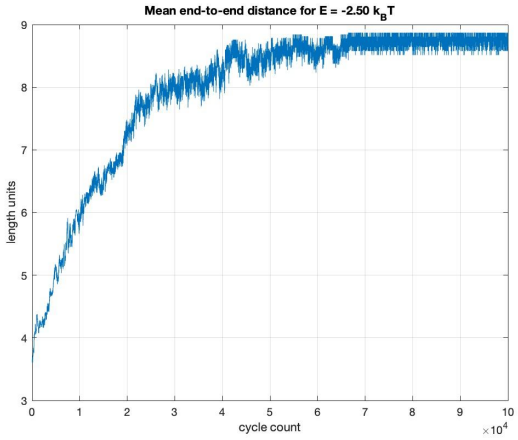
The gyration radius plot does indicate the slower increase in size for the larger negative values of interactions energy per non covalent bond. The bottom right plot for -2.50 kT units reaches the value near 8 kT of total energy much slower than subsequent lower magnitude plots. This is consistent with the results from the energy values.

Mean end-to-end distance plots for each cycle in the 50 repetitions:









Analysis of the end-to-end distance of the protein model:

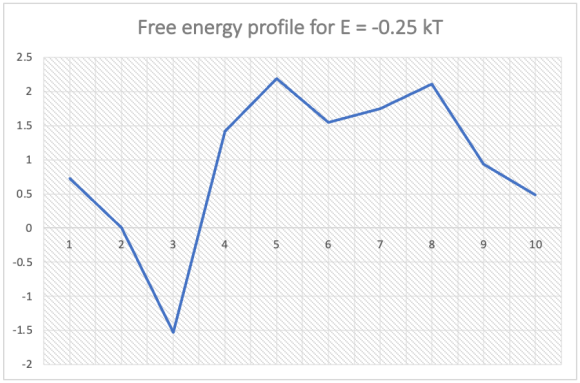
While collecting data for this parameter, the cartesian distance between the 1st and the 16th residues were calculated for every cycle. This data has also been subjected to the same averaging as in the previous two datasets.

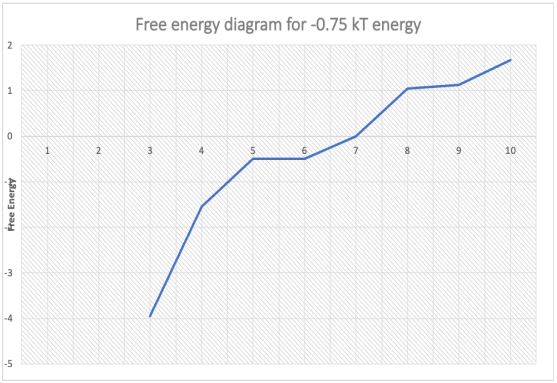
This parameter is also helpful in analysing the unfolding of the protein as it shows the distance of separation between the two ends. However, this cannot be regarded as a direct measure of the size of the protein for reasons as stated before. A quick analysis of the plots shows that the end-to-end distance for more negative values of the interaction energy per non-covalent bond length increases slower than those for the less-negative values.

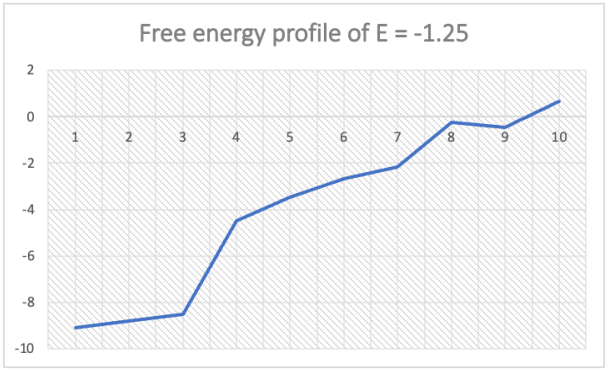
The mean end-to-end distance increases very slowly in the 4th case indicating the very slow unfolding of protein in the case of high interaction energies.

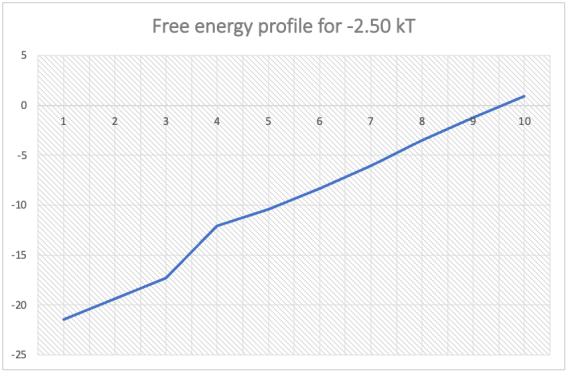
This is also consistent with the hypotheses that have been obtained in the analysis of the first two parameters.

Constructing a free energy profile:









The above free energy profiles are constructed by counting the number of microstates using the energy data. The no of microstates were used to find average frequencies which were used to calculate the entropy of each state. The entropy was then used to calculate free energy. The graph is as shown above.

Number of microstates sampled:
This can essentially be considered as the number of unique gyration radius values per energy level in the whole data set.

Setting threshold for unfolding:
Observing the plots for radius of gyration for all the energy values, the unfolded protein molecule never attains a mean gyration radius below *6 length units*. Hence, we can say that the protein is completely unfolded when it first reaches the radius of gyration of 6 length units.
For different levels, the number of cycles required to get to *6 length units* is as follows:

Energy level (in kT units)	No of microstates	No. of cycles required on an average to unfold
-2.50	7622	21712
-1.25	8635	5795
-0.75	11141	2924
-0.25	12565	2475

Analysing the no. of cycles required to unfold
The number of cycles taken for the molecule to unfold is clearly increasing with increase in the interaction energy per bond. This is because of the fact that molecules will find it harder to unfold when the interaction between residues is stronger than when the molecule has smaller inter-residue interactions.

The two factors at play in this are the possible complex free energy landscape of the molecule and the metropolis criterion. The nature of the metropolis criterion is such that it favours the states which reduce the overall energy of the model. The free energy landscape of the molecule may tend towards a gaussian-like state due to the central limit theorem, where the complex can be said to require a given number of cycles to escape a minimum. Since this number can be estimated to be a gaussian distribution, the sum is also gaussian.

When the interactions are considered between all proximal residues:
The number of cycles it will take to unfold the protein is much slower than the number of cycles required for the case of initially defined interactions. This is because of the fact that the protein will find newer folded conformations when it samples different microstates throughout its random walk.

We considered 10000 cycles and 20 iterations. For this we obtained the mean no of cycles at which the molecule unfolds (considering the radius of gyration threshold of unfolding as 6 length units) to be **2534.** Although a lot of variation was observed in this data and this value has high error.