

# Simulation of Folding for a Granular Chain in Water

Ben Ball<sup>a</sup>

<sup>a</sup>*Department of Physics, Applied Physics, and Astronomy,  
Rensselaer Polytechnic Institute, Troy, NY 12180*

**Abstract.** One novel approach to polymer folding is the use of macroscopic physical models. A previous experiment used a metal chain in water to imitate the polymer folding process [1,2]. A simulation of this experiment is run in an attempt to recreate similar results. Written in Interactive Data Language, this simulation is based on a square lattice model [3-7], where every particle has an integer position on a two dimensional grid. The simulation is run through a series of steps in which one of many possible moves is selected, with a probability based on the potential energy change in which that move results. Many parameters of the simulation are varied, including the chain length, the potential it experiences, and the restrictions on how it can move. The resulting data has many similarities to that of the physical chain experiment, and shows relationships also observed in polymer folding [7].

## INTRODUCTION

The study of how molecular polymer chains collapse in poor solvents is an ongoing field of research. Not only is polymer collapse related to the early stages of protein folding, it also allows us to investigate many other biological molecules such as DNA [1]. A recent experiment attempted to study polymer folding using a macroscopic chain. A steel chain made of loosely connected spheres is placed in a thin layer of water, and shaken vertically [1-2]. The chain folds in upon itself because of an attractive force from the surface tension. A simulation is used to verify the results of this experiment, and explore more possibilities which are not possible in a physical experiment. Looking at several polymer and protein folding simulations as an example, it was decided that this simulation would use a square lattice method [3-7].

## MODEL

A computer program is used to simulate the prior experiment. Rather than calculate the forces and accelerations of each particle, a statistical model is used, where the energy of different possible states is evaluated and used to create a probability distribution of moves. The vertical movement of the plate and subsequent Brownian movement of the water are not directly implemented. Rather, these are assumed to contribute to an equivalent “temperature” which provides energy for the system to move.

## Program Description

The entire computer simulation is done on a two dimensional square lattice, similar to those used in polymer folding simulations [3-7]. It takes a starting configuration of the chain (usually a circle with the two ends slightly separated) and a set of parameters, and repeatedly performs a step operation, usually between 500 and 2000 times for the results presented here. The step operation can be broken down into two parts. First the possible moves are found and evaluated to find the energy cost associated with them, and then a move is selected and made.

### *Evaluating the Moves*

From each monomer in the chain, 14 different moves, which will be described later, are considered. For each of these moves, there are two parts. First, the new position generated by this move is examined to discover if it violates any of the simulation rules. The rules are that only one monomer can be at each location on the lattice,

adjacent monomers in the chain are not further than a maximum link length from each other, the chain cannot cross itself, and two rules which constrain the angles that can be formed by the chain. The first is a maximum angle between each set of three monomers, measured as the angle away from a straight line. The second is a limitation that two of these angles next to one another cannot add up to more than 180 degrees, preventing tight U turns. This last rule is turned off for some simulations.

Second, the total energy of the new position is calculated. There are three components to this energy. The first is surface tension energy, which is proportional to the distance between two monomers, summed over all possible pairs ( $(L*(L-1))/2$  in total). The second is the link potential between adjacent monomers, and is proportional to the offset from an equilibrium distance between each pair of monomers ( $L - 1$  in total). Finally, the third potential is proportional to the angle between each set of 3 adjacent monomers ( $L - 2$  in total). For all three potentials, both a leading coefficient and the power of the relationship can be changed.

### *Types of Move*

The 14 different moves are divided into three groups. The first 8 moves consist of moving the monomer a single step in each direction on the grid, including diagonals. The next two moves are considerably more complex, and shift up to half the chain at once. The starting monomer is shifted one step in either direction (thus two moves). The directions are determined as perpendicular to the line formed by the starting monomer and the monomer adjacent to it which remains stationary. The portion of the chain which will remain stationary is always the longer half. Each monomer further along the chain is also shifted one step in the same direction, relative to the line formed by its previous position and the previous position of the monomer behind it. However, if one of these movements would cause a rule to be broken, the monomer is allowed to shift in a different direction or remain still. First the two directions adjacent to the initial one are tried, then the next two further away, until all eight have been tried, then the current position (no move for that monomer) is tried. The first position which does not break any rules is used, and then the next monomer down the chain is placed. If at any point all 9 possible moves are rejected, the entire shift operation is rejected. Once the last monomer of the chain is placed, the positions are final and the energy is evaluated. At no point does the process rewind to try a different move at a previous monomer.

The third and final group of 4 moves is very similar to the previous group. Instead of moving the shorter portion of the chain, either half can be moved (therefore creating 4 moves). Also, when trying to place each monomer after the first, the first position tried is the current position (no move), then the position in the same direction as the initial move. This often results in a middle section of the chain shifting while both ends remain stationary.

### *Choosing a Move*

At this point, the simulation has a list of all the moves which were successful and the energy of the resulting state. The probability of each move is calculated according to the formula  $P = \exp(\Delta U/T)$  where  $\Delta U$  is the change in energy due to the move, and  $T$  is a parameter of the simulation [7]. This formula is adapted from a molecular polymer folding lattice model. These probabilities are then normalized, and a move is chosen at random. The state of the chain is updated accordingly, and the simulation moves on to the next step.

## **Parameter Explanation**

There are many places throughout the program where a specific value was chosen for use in an equation or procedure. These chosen values are summarized in Table 1. It is important to note that these values were not chosen arbitrarily. A brief explanation of each choice follows.

The first set of parameters are the Coefficients and powers of the potentials. The absolute magnitude of the three potential constants has no effect on the simulation; only their relative magnitudes are important. The values shown here were selected to give the potentials approximately equal weight. The angle potential is significantly smaller because it is multiplied by an angle measured in degrees, a significantly larger number than the distances used for the other two potentials. The surface tension potential includes dependence on the square of the length of the chain. This is because the surface tension scales with the square of the length, unlike the other two which are linear, because it is counted for each pair. It also scales with length because longer chains tend to have longer distances between monomers.

The power of the surface tension potential is one, yielding a constant force. This is physically accurate: surface tension provides a constant force independent of distance [8]. The link potential uses a quadratic potential,

simulating a spring. Finally, the angle potential was arbitrarily chosen to be linear. A quadratic potential was tested, but no significant difference was observed.

The chain has a circular initial position. This is to mimic the experiment, and to allow easier folding. A maximum angle less than 180 was tested, but found too restrictive to the folding process. However, to prevent sharp turns and force the formation of pearls similar to those in the experiment, the simulation was not allowed to have to adjacent angles that sum to 180 degrees or more. The link equilibrium length was chosen to be 1 because this is the base spacing of the monomers on the grid. However, a maximum link length of 1.5 allows far more freedom of movement, because monomers can be on diagonal vertices, rather than only directly adjacent to their neighbors. Finally, the temperature was chosen as a medium value, allowing some freedom in the movement of the chain without making it to random. It is dependent on the length of the chain because it is directly related to the energy, which also scales with the length of the chain.

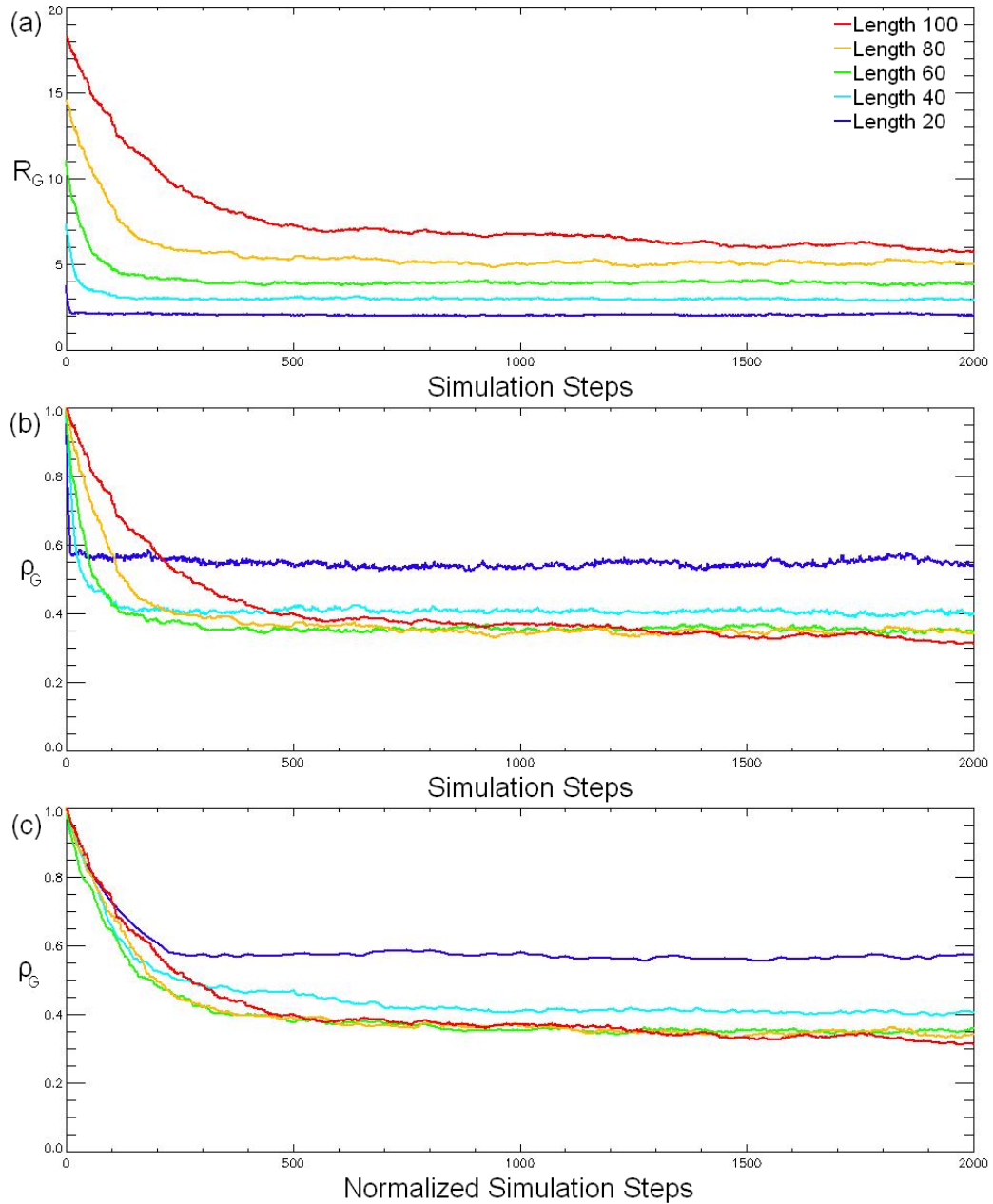
**TABLE (1).** Unless otherwise noted, all parameters for the data presented here use the below default values.

| Parameter                                                 | Default Value                        |
|-----------------------------------------------------------|--------------------------------------|
| Coefficient of surface tension potential                  | $8000/(L^2)$ where L is chain length |
| Power of surface tension potential                        | 1                                    |
| Coefficient of link potential                             | 100                                  |
| Power of link potential                                   | 2                                    |
| Coefficient of angle potential                            | 3                                    |
| Power of angle potential                                  | 1                                    |
| Starting position                                         | Approximate circle                   |
| Maximum angle                                             | 180 (equivalent to no limit)         |
| Limit that two adjacent angles cannot sum to 180° or more | Present                              |
| Link equilibrium length                                   | 1                                    |
| Link maximum length                                       | 1.5                                  |
| T                                                         | $5*L$                                |
| Moves done                                                | All 3 types                          |

## RESULTS AND ANALYSIS

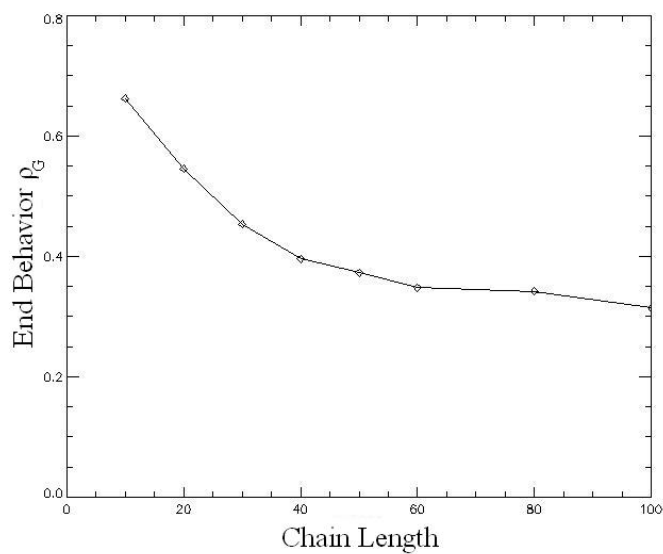
To facilitate a quantitative analysis of the folding of the chain, the radius of gyration ( $R_G$ ) is observed. This is simply the root-mean-square distance of all the particles from the center of mass. Figure 1a shows a graph of  $R_G$  as the simulation progresses, for five different lengths of chain, 20, 40, 60, 80 and 100 particles. Each line is the average of five trials with the same parameters. The longer chains show longer both starting and ending  $R_G$  values, as one would expect because the starting circle must have a larger circumference and therefore radius, and the larger number of monomers will never be able to compact into as small a space. Therefore to gather meaningful data across multiple lengths, a normalization based on length of chain must be used. Figure 1b shows the same graph, with each data set divided by the initial value. Because the starting configuration is the same for all simulations, this is equivalent to dividing by the length of the chain, but is appealing because the starting value is now 1. This normalized  $R_G$  is called  $\rho_G$  for consistency with the experiment, all though it is calculated differently. In the results of the physical model, the  $R_G$  values are multiplied by  $2\pi/L$ . Since the initial formation of the chain is a circle and therefore all the points are equidistant from the center, the  $R_G$  is simply the radius of the circle. The circumference of the circle must be L, and therefore this conversion gives a starting  $R_G$  value of  $2\pi * \text{radius}/\text{circumference}$ , which is always one. Therefore the conversions done on both sets of data are equivalent. After these changes, Figure 1b still shows significant difference in the folding rates of the different length chains. A comparison of the folding times for the different lengths led to the determination that an appropriate normalization factor is  $L^2$ . Thus, the x-axis of each length is multiplied by  $10,000/L^2$ . This leaves the time scale for the length 100 chain the same, and spaces out the other data so that it all shows the same basic timescale. This time-normalized version can be seen in Figure 1c.

In a detailed simulation of polymer folding, the folding speed was found to be proportional to  $L^{1.74 \pm 0.08}$  when the surrounding solvent moves with simple Brownian motion, the situation most similar to this simulation [9]. Despite the difference in method of simulation and also the type of folding event being simulated, it is interesting that a similar length dependence appears in both simulations. The length dependence for this simulation was not explicitly calculated due to uncertainty as to when the chain has completely folded, and therefore the actual power of the relation is most likely above or below the estimated value of 2.

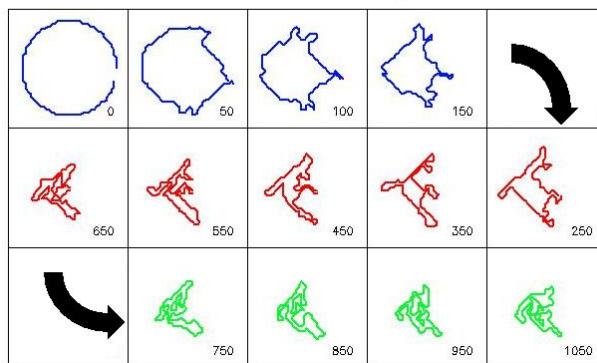
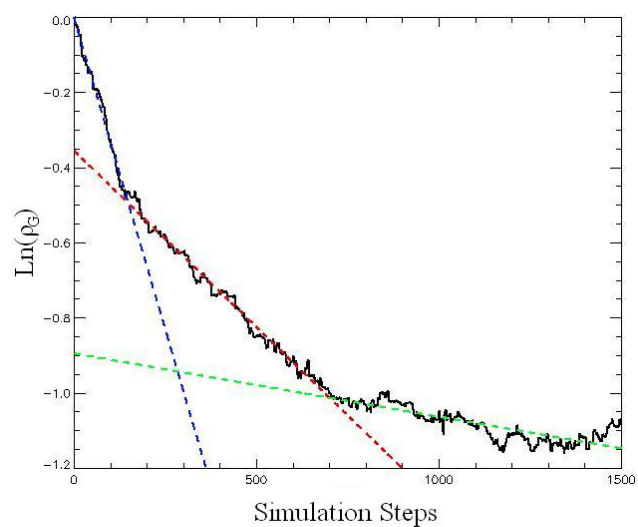


**FIGURE 1.** A comparison of simulation data with and without normalization. Figures (b) and (c) have the radius of gyration normalized by dividing by the initial value, and (c) has the time normalized by the square of the chain length. All results here are the average of 5 trials.

Also of interest is the relationship between  $\rho_G$  and the length of the chain. While Figure 1c has been normalized with respect to length, there is still a clear difference in the long timescale minimum reached by the different lengths. Figure 2 shows a plot of this long timescale minimum  $\rho_G$  value as a function of length, with data points at 10, 20, 30, 40, 50, 60, 80 and 100. Once again each is the average of five runs with the same parameters. Note the non-linear relationship, and higher  $\rho_G$  values at shorter chain lengths. The experiment conducted previously also found a similar relationship, with  $\rho_G$  decreasing as the chain length increases [1].



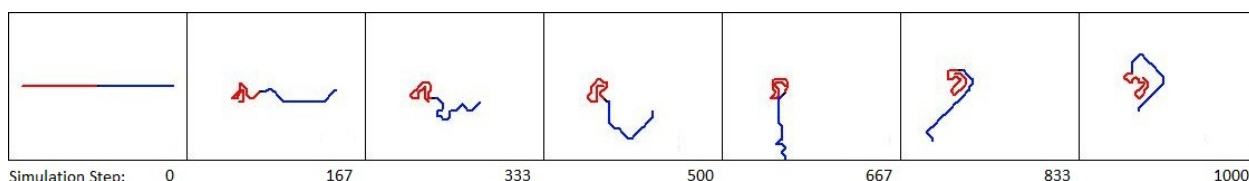
**FIGURE 2.** A graph of  $\rho_G$  values after 2000 simulation steps, as a function of chain length. Averaged over 5 runs. Note the downward, nonlinear trend.



**FIGURE 3.** A graph of  $\text{Ln}(\rho_G)$  as a function of simulation step, for a chain which has a potential link between its ends. Observe the three timescales, corresponding to the three colors in the second image. Note that the first 4 frames of the image are separated by steps of 50, and the next 9 are separated by steps of 100.

In Figure 1,  $\rho_G$  appears to be a uniform curve decreasing in slope as it approaches a minimum value. However, when the chain is closed by adding a high potential attraction between the two ends, a more interesting graph is observed. As shown in Figure 3, the log graph of  $\rho_G$  versus simulation step shows three distinct linear regions. Since this is a log plot, each of these regions corresponds to a different polynomial fit. The highest (negative) degree fit is first, and the lowest degree fit is last.

A similar plot showing three timescales was also found for the physical experiment [1]. The proposed explanation for the three modes of collapse in this model is that there are three different physical procedures occurring: first, the chain forms loops, or pearls, then these pearls compete between one another, and finally the chain compacts without changing its overall shape. This explanation also fits the results presented here, to a limited extent. By the first red image in Figure 3, four pearls have formed. Over the red images these pearls absorb the rest of the chain and compete with one another, and by the first green image, only three pearls remain. Finally over the green images the chain compacts while retaining the three existing pearls. In the data presented here, the ratio of the slopes between the three timescales is 19.7 to 5.6 to 1, from left to right. In the physical experiment, the ratio of timescales found was 15 to 5 to 1. The similarity between the ratios in experiment and simulation supports the possibility that the same phenomenon is being observed in both cases.



**FIGURE 4.** A diagram of the position of a chain over 1000 simulation steps. The red half of the chain has normal properties, while the blue half has no surface tension potential.

An experiment done with the physical model involved coating half of the chain in surfactant which changes the surface tension interactions with the water. This caused that half of the chain to no longer experience a drive to fold. In that experiment, half the chain was observed to remain nearly linear while the other half of the chain folded in on itself. After some time, the surfactant coated half wrapped around the already folded half of the chain [1]. As shown in Figure 4, a very similar effect was created in the simulation. This was done by eliminating the surface tension energy from half of the chain.

## CONCLUSIONS

The intent in the creation of this model was to accurately simulate the previous granular chain folding experiment. The two models have similar plots of radius of gyration, and show similar dependence on length of chain. In both cases three different folding timescales are observed, and the relative rates of these timescales are similar. When the simulation is varied by removing the driving force of collapse from half of the chain, the two models show very similar movement of the chain. While differences between the models exist and not all data agrees, clearly this goal has met with a significant amount of success. The next step is to use this simulation to study that data which is difficult to gather in the physical experiment, and to compare the results from these models to the greater polymer folding community. This has already been started, with the comparison of the dependence of folding time on chain length, but there are many more areas that still need further investigation.

## ACKNOWLEDGMENTS

The author would like to acknowledge support in funds by the CASPER program and the NSF REU program. The author would also like to acknowledge the support in time, direction, and assistance by Jeffrey Olafsen, who advised the author in the completion of this project.

## REFERENCES

1. B. Bammes and J. S. Olafsen, Preprint.

2. B. Bammes and J. S. Olafsen, *Chaos* **14**, S9 (2004).
3. R. Mélin, H. Li, N.S. Wingreen, and C. Tang, *J. Chem. Phys.* **110**, 1252 (1999).
4. A. Šali, E. Shakhovich and M Karplus, *J Mol Biol* **235**, 1614-1636 (1994).
5. Z Li and H. A. Scheraga, *Proc Natl Acad Sci U S A.* **84**, 6611-6615 (1987).
6. V. I. Abkevich, A. M. Gutin, and E. I. Shakhnovich, *J. Chem. Phys.* **101**, 6052 (1994).
7. K. F. Lau and K. A. Dill, *Macromolecules* **22**, 3986-3997 (1989).
8. R. Cini, G. Loglio, and A. Ficalbi, *J. Colloid Interface Sci.* **41**, 287-297 (1972).
9. N. Kikuchi, A. Gent, and J.M. Yeomans. *Eur. Phys. J. E* **9**, 63-66 (2002).