Bootcamp Project: Polymer Chain Conformation under Different Solvent Conditions

The Liang Lab

October 2025

Goals

1. Perform molecular dynamics (MD) simulations of a coarse-grained polymer chain in a dilute solution using LAMMPS.

2. Analyze simulation data to quantify structural properties of polymer chains, such as radius of gyration and end-to-end distance.

3. Investigate how solvent quality influences polymer chain conformation.

Instructions

Step 1: Learn LAMMPS

LAMMPS is an open source classical MD code. The best way to learn LAMMPS is to read its manual (https://docs.lammps.org/Manual.html), install LAMMPS on your computer, and try to run simple simulations. In this bootcamp project, you will run simulations of a bead-spring polymer chain in an implicit solvent. Following is an example of LAMMPS input script, in.run (we often name the input script as in.xxx), for the simulation.

```
1  # An example LAMMPS input script
2  variable        eps equal 0.1
3  variable        rcut equal 2.5
4  variable        T equal 1.0
5  variable        damping equal 10.0
6
7  units           lj
8  atom_style      molecular
9  dimension       3
10 boundary        p p p
11 neighbor        1.5 bin
12 neigh_modify    every 1 delay 0 check yes
13
14 read_data       data.chain
15 velocity        all create ${T} 4928459 rot yes dist gaussian
16
17 pair_style      lj/cut ${rcut}
18 pair_modify     shift yes
```

```
19  pair_coeff      1 1 ${eps} 1.0 ${rcut}
20
21  bond_style      fene
22  bond_coeff      1 30.0 1.5 1.0 1.0
23
24  special_bonds   fene
25
26  fix             1 all nve
27  fix             2 all langevin ${T} ${T} ${damping} 48279
28  timestep        0.01
29
30  compute         rg all gyration
31  fix             3 all ave/time 1000 1 1000 c_rg file gyration.dat mode scalar
32
33  dump            1 all custom 10000 dump.chain id mol type x y z ix iy iz
34  dump_modify     1 flush yes
35
36  thermo          1000
37  thermo_style    one
38  thermo_modify   flush yes
39
40  run             10000000
41
42  write_data      equil.data
```

Listing 1: An example of LAMMPS input script: in.run

Step 2: A Simple yet Powerful Model of Polymer: The Bead-Spring Model

The bead-spring model is the simplest model that captures the universal behavior of polymers. This model is introduced by Kurt Kremer and Gary Grest in their seminal paper *Dynamics of entangled linear polymer melts: A molecular-dynamics simulation.* (Kremer, K. and Grest, G.S., *J. Chem. Phys.* **1990**, 92(8), 5057-5086). In this model, many spherical beads (i.e., *monomers*) are connected by springs (i.e., *bonds*) to form a polymer chain. Non-bonded beads interact with each other through a truncated-shifted Lennard-Jones potential (defined in lines 17-19 in in.run. Bonded potential is discribed by the finite extensible nonlinear elastic (FENE) potential (defined in lines 21-22 in in.run).

Questions to Think About (Step 2)

Read LAMMPS manual and answer the following questions:

1. How does the Lennard-Jones potential change with different interaction strength (`${eps}` in in.run) and cutoff distance (`${rcut}` in in.run)? Why is it truncated and shifted in simulations?

2. Given the `bond_coeff` defined in in.run, what would be the equilibrium bond length?

3. What does `special_bonds fene` mean in in.run (at line 24)? Why is this setting important?

Step 3: A Basic MD Simulation

Understand the basics of a MD simulation.

- **Reduced units:** Instead of SI units, the coarse-grained model uses reduced units. The units of length, energy, and mass are σ, ϵ, and m, respectively. And the Boltzmann constant $k_B = 1$.

- **Periodic boundary condition (PBC):** The PBC is used by almost all MD simulations. By applying the PBC, we can model the behavior of bulk materials while simulating only a small representative region of the system. The PBC is set at line 10 in in.run.

- **Explicit solvent vs implicit solvent:** In explicit solvent simulations, the solvent molecules are represented explicitly by individual solvent particles. In contrast, implicit solvent simulations do not include solvent particles; instead, the solvent's influence is incorporated through effective forces and thermostats.

  In implicit solvent simulations, the effect of solvent is modeled by a Langevin thermostat (line 27 in in.run), where the damping parameter (`${damping}` in in.run) captures the viscosity of the solvent. The solvent quality is captured by the interaction strength between solute beads (i.e., monomers on the polymer chain).

- **Initialization:** The simulation box size, beads coordinates and bonded topology are initialized by reading an LAMMPS data file (line 14 in in.run). The velocities of beads are initialized at line 15 in in.run. Read LAMMPS manual (https://docs.lammps.org/read_data.html) for the format of LAMMPS data file. Following is an example LAMMPS data file of a single chain with $N = 5$ monomers:

```
1   LAMMPS data file for a bead-spring chain with N=5 beads
2
3   5 atoms
4   1 atom types
5
6   4 bonds
7   1 bond types
8
9   -5.0 5.0 xlo xhi
10  -5.0 5.0 ylo yhi
11  -5.0 5.0 zlo zhi
12
13  Masses
14
15  1 1.0
16
17  Atoms # molecular
18
19  1 1 1 0.0 0.0 0.0
20  2 1 1 0.97 0.0 0.0
21  3 1 1 1.94 0.0 0.0
22  4 1 1 2.91 0.0 0.0
23  5 1 1 3.88 0.0 0.0
24
25  Bonds
26
27  1 1 1 2
28  2 1 2 3
29  3 1 3 4
30  4 1 4 5
```

Listing 2: An example of LAMMPS input data file: data.chain.N5

The chain in the example data file is too short. Use python to generate a data file of a polymer chain with $N = 100$. Remember to change the box dimension (see Question 2)

- **Integrator:** LAMMPS uses velocity-Verlet integrator, which is a Symplectic integrator. The integrator is defined at line 26 in in.run. The integration timestep is defined at line 28 in in.run.

Questions to Think About (Step 3)

1. Why use reduced units in coarse-grained simulations? How to map back to real units?

2. To simulate a polymer chain in a dilute solution, what is the minimum dimension of the simulation box?

Step 4: Run a simulation

A standard simulation includes the following steps:

1. **Initialization:** Explained in Step 3.

2. **Relaxation:** The initial configuration is usually not in **thermodynamic equilibrium**, while most material properties are defined under equilibrium conditions. A relaxation run is therefore required to bring the system to equilibrium.

3. **Production:** After equilibration, a production run is carried out to collect sufficient simulation data for evaluating material properties. Simulation trajectory provides a lot useful information, and is saved by `dump` command in LAMMPS (lines 33-34 in in.run).

Step 5: Visualization and Analysis

Visualization

One advantage of computer simulations over experiments is that you can directly observe the motion of individual molecules. To do so, you can visualize the simulation trajectory (`dump.chain`) by OVITO (https://www.ovito.org/) or VMD (https://www.ks.uiuc.edu/Research/vmd/).

Analysis

Analyze the chain conformation from the simulation trajectory of the production run.
   Mean-squared radius of gyration:

$$\left\langle R_g^2 \right\rangle = \left\langle \frac{\sum_{i=1}^{N} m_i \left(\vec{r}_i - \vec{r}_{com} \right)^2}{\sum_{i=1}^{N} m_i} \right\rangle \tag{1}$$

where $\vec{r}_{com}$ is the center of mass of the chain, $\vec{r}_i$ and m_i are the coordinate and mass of i-th bead, $\langle ... \rangle$ represents a time average.
   Mean-squared internal distance is the mean-squared distance between two monomers separated by n bonds on the same chain:

$$\left\langle R^2(n) \right\rangle = \left\langle \left(\vec{r}_{i+n} - \vec{r}_i \right)^2 \right\rangle \tag{2}$$

where $\langle ... \rangle$ denotes both a time average and an average along the chain (e.g. $i = [1, N - n]$).

Questions to Think About (Steps 4 and 5)

1. How does LAMMPS record the simulation trajectory? What are `x`, `xs`, `xu`, `xsu` coordinates in LAMMPS dump file? What is `ix`? If a chain crosses the periodic boundary, how to treat its coordinates?

2. The initial conformation of the chain can be either a straight line (as in the example data.chain.N5) or a random coil. How does the initial conformation influence the simulation—for example, the number of steps required for relaxation or the equilibrium properties such as the radius of gyration? Use python to generate different initial conformations and test your hypothesis.

3. How long should a relaxation run be, and how can we determine whether the simulated system has reached equilibrium?

4. How to estimate error bars of $\langle R_g^2 \rangle$ and $\langle R^2(n) \rangle$? (Hint: block averaging)

5. How can solvent quality be inferred from chain conformation? (Hint: Plot $\langle R^2(n) \rangle$ vs n in log-log scale and fit the slope.)

Step 6: Run the Simulation under Different Conditions

Vary the pairwise interaction strength (change `${eps}` in in.run from 0.1 to 1.0), and repeat Steps 4 and 5. Determine the conditions corresponding to good, theta, and poor solvents. Compare your results with the literature report (Graessley, W. W., Hayward, R. C., and Grest, G. S., *Macromolecules*, **1999**, 32, 10, 3510–3517).

Questions to Think About (Step 6)

1. Will $R^2(n)$ depends on the degree of polymerization of the chain, N? If yes, why? Try to run simulations with different N under the same good solvent condition and verify your hypothesis.

2. Will the cutoff distance (`${rcut}` in in.run) of the Lennard-Jones pairwise potential affect solvent quality? Will the tail correction affect chain conformation? Run simulations to test your hypothesis.

Step 7: Report the Results

Great work! You've made it to the end of this bootcamp!

You've learned how to use LAMMPS to simulate a polymer chain in implicit solvent and how to analyze its conformation from the resulting trajectory. The final step is to communicate your results effectively, just as you would at a scientific conference in front of colleagues and experts.

Below are guidelines to help you create clear and professional slides:

1. **Start With an Introduction**

   - 1-2 slides of background information (i.e., chain statistics in different solvents)
   - Clearly state the scientific question. What are you studying, and why does it matter?

2. **Describe the Method Clearly**

- Includes essential details of your simulation model so that others can reproduce it, such as
  - Force fields and parameters used
  - System size, box size
  - Temperature and pressure control
  - Integration algorithm, timestep
- Provide a schematic of your model. A clear simulation snapshot works great.

3. **Summarize Your Analysis and Highlight Key Findings**

- Show how you analyze the simulation trajectory. How are chain conformations characterized? ($\langle R^2(n) \rangle$, power-law fitting)
- Provide plots with readable labels and captions.
- Summarize your results. How does the chain conformation change with the interaction potential?
- *Convey insights, not just numbers.* Do your results agree with theoretical predictions, computational results from other references, and/or experimental observations? Why or why not? What is the molecular-level/microscopic mechanism?

4. **Conclude With Take-Home Messages**

- Summarize what you learned about polymer conformations in implicit solvent.
- Discuss limitations of the simplified model used in this project.
- Suggest possible next steps or follow-up simulations.

5. **Tips**

- Avoid long paragraphs; rely on bullet points.
- Use large figures with clear labels that are readable from a distance.
- Use figures/pictures when possible. A picture is worth a thousand words.
- Maintain consistent fonts, colors, and spacing throughout the presentation.
- For a 10-minute conference-style talk, aim for approximately 8-12 slides.