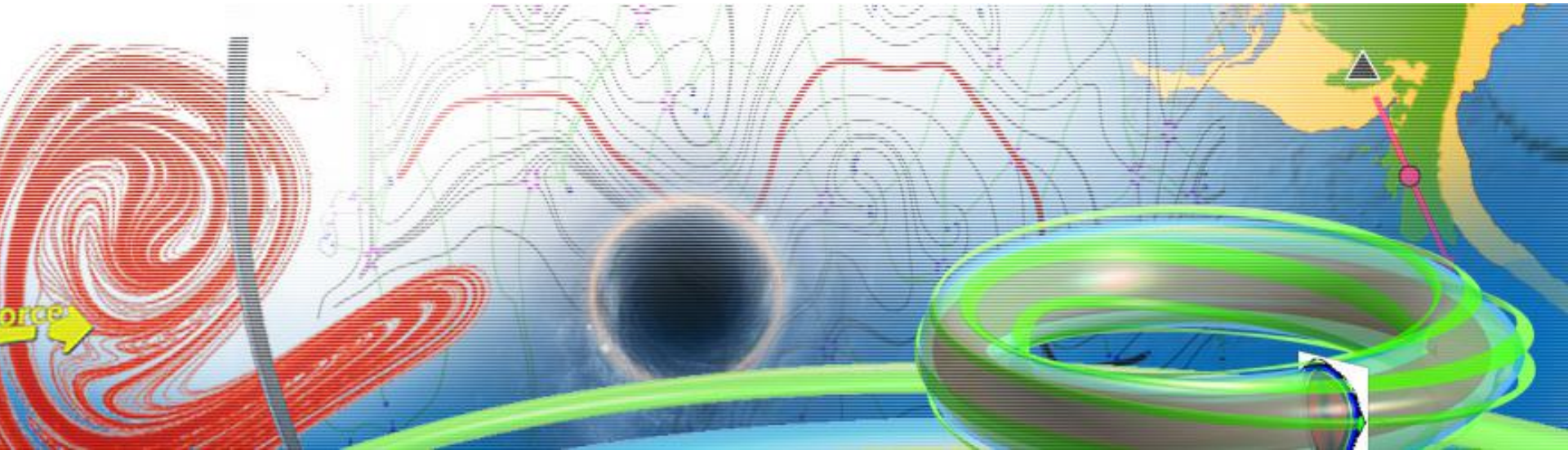
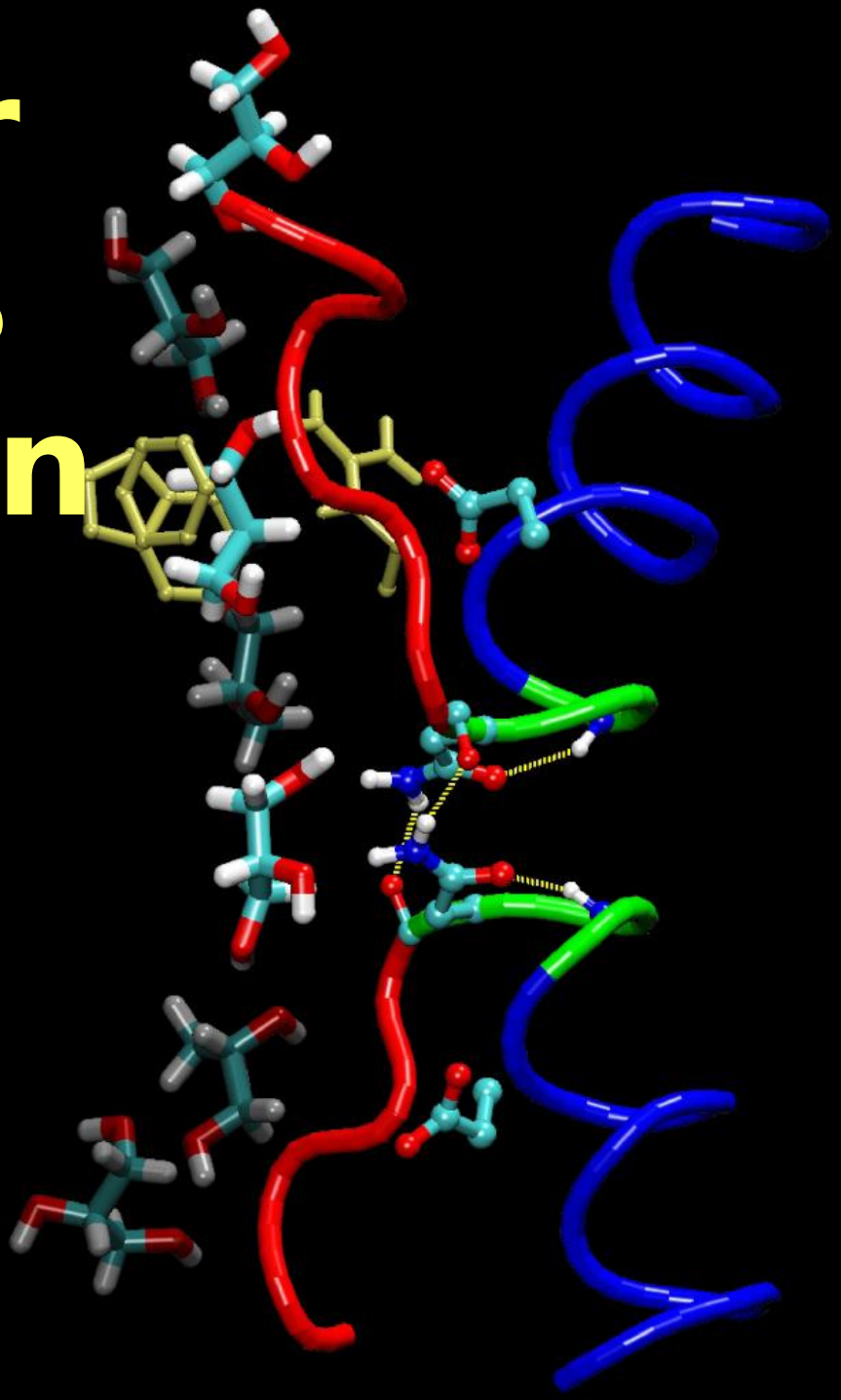
生物动力系统模拟

王冠宇 18665955633
wanggy@sustc.edu.cn

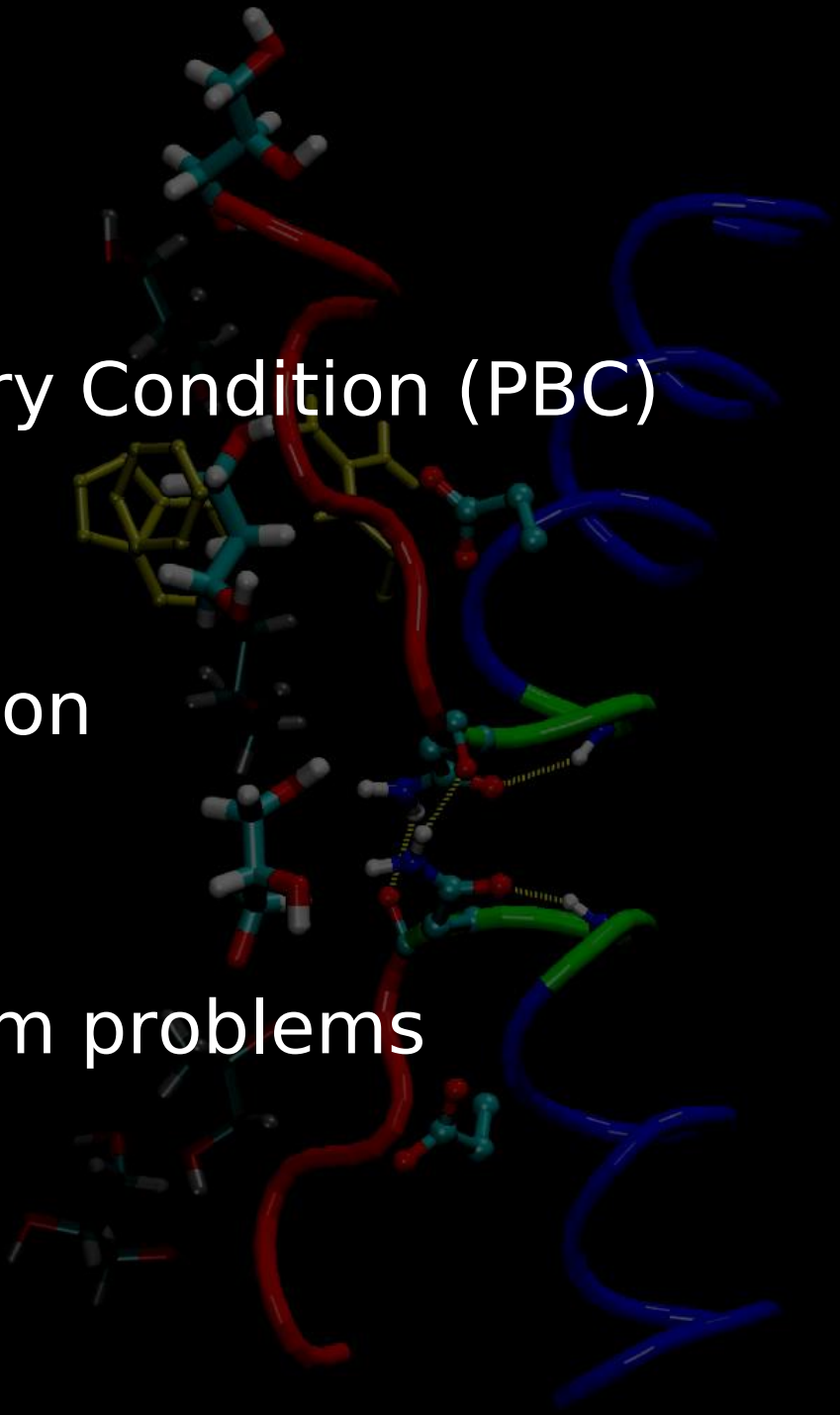


Molecular Dynamics Simulation



Contents:

- Periodic Boundary Condition (PBC)
- Steps of simulation
- Middle term exam problems



6. To quit xLEaP use **quit.**

```
> quit
```

All the above steps are just for preparati
Now let’s get into the real meat.

1. Minimization

2. Heating with constant volume and temperature (NVT) for 20ps from 0K to 300K

3. Production MD with constant pressure and temperature (NPT) at 300K and 1atm for 60ps

Minimization

Why minimization?

An initial structure is likely to be **NOT** in its most natural structure

This can be easily seen from the simplest structure: a bond

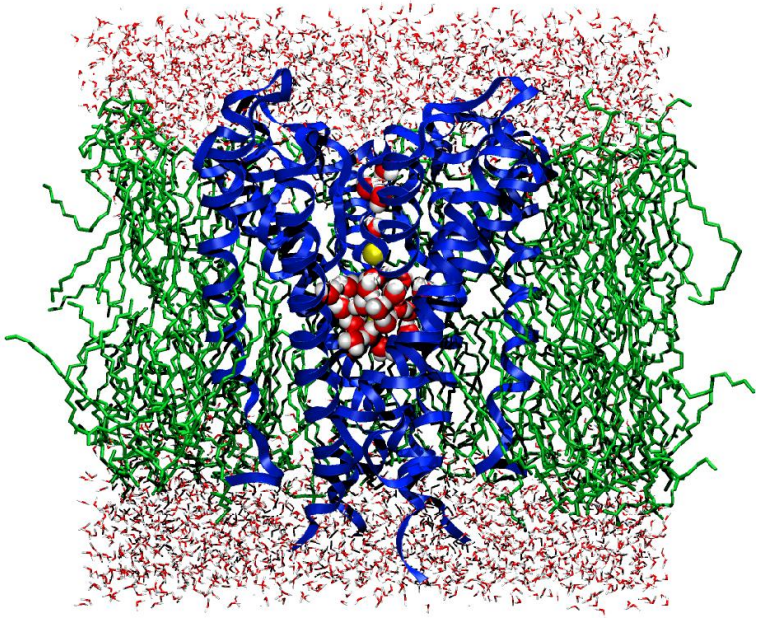
$$V = \frac{1}{2}k(r - r_0)^2$$ if other forces are ignored

As long as $r \neq r_0$, it is not natural

Therefore, the natural state corresponds to $V = 0$, energy minimal.

To begin MD with the most natural structure, one must perform min

But the real scenario is not that simple



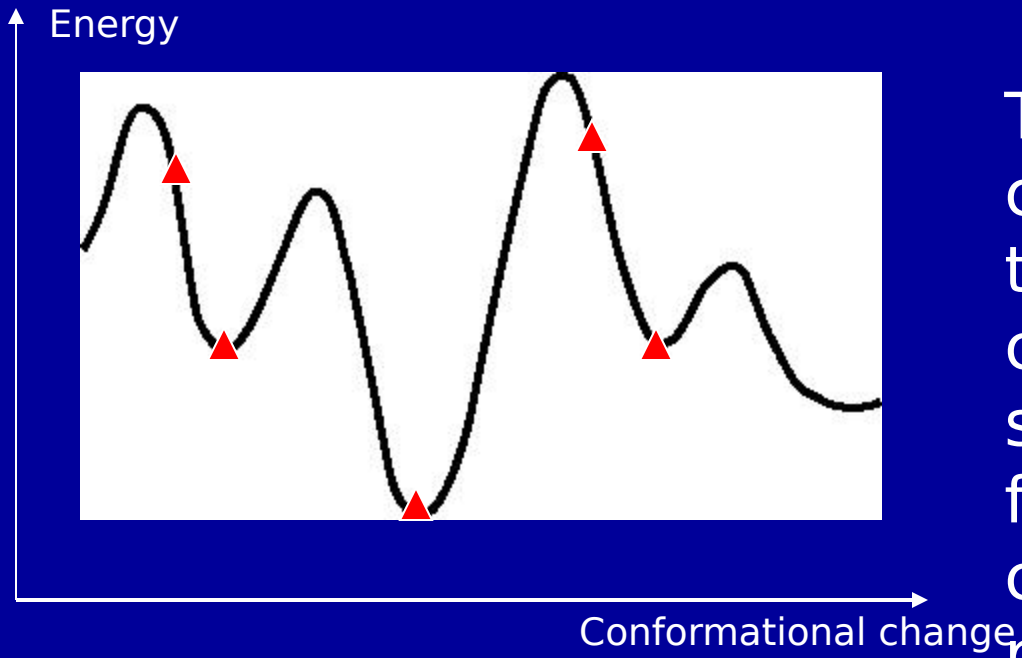
$$V_{\mathrm{AMBER}} = \sum_{i}^{n_{\mathrm{bonds}}} b_i(r_i - r_{i,\mathrm{eq}})^2 + \sum_{i}^{n_{\mathrm{angles}}} a_i(\theta_i - \theta_{i,\mathrm{eq}})^2 + \sum_{i}^{n_{\mathrm{dihedrals}}} \sum_{n}^{n_{i,\mathrm{max}}} (V_{i,n}/2)[1 + \cos(n\phi_i - \gamma_{i,n})]$$
$$+ \sum_{i<j}^{n_{\mathrm{atoms}}}{}' \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + \sum_{i<j}^{n_{\mathrm{atoms}}}{}' \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}},$$

1. **The terms are coupled, reduce one would increase anoth**

2. **The equation is too complex to minimize analytically**

3. **Numerical minimization is required.**

Numerical methods for minimization

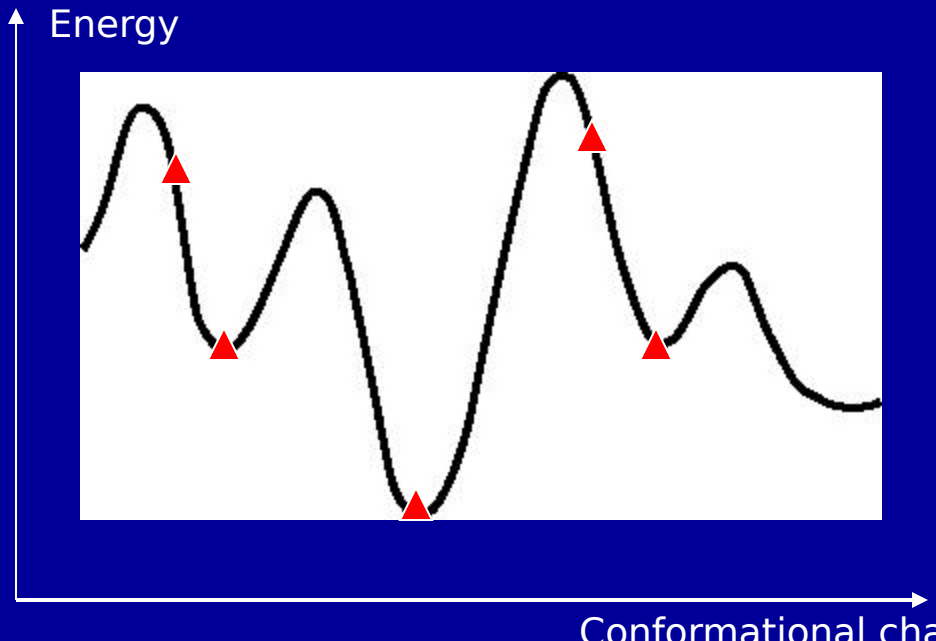


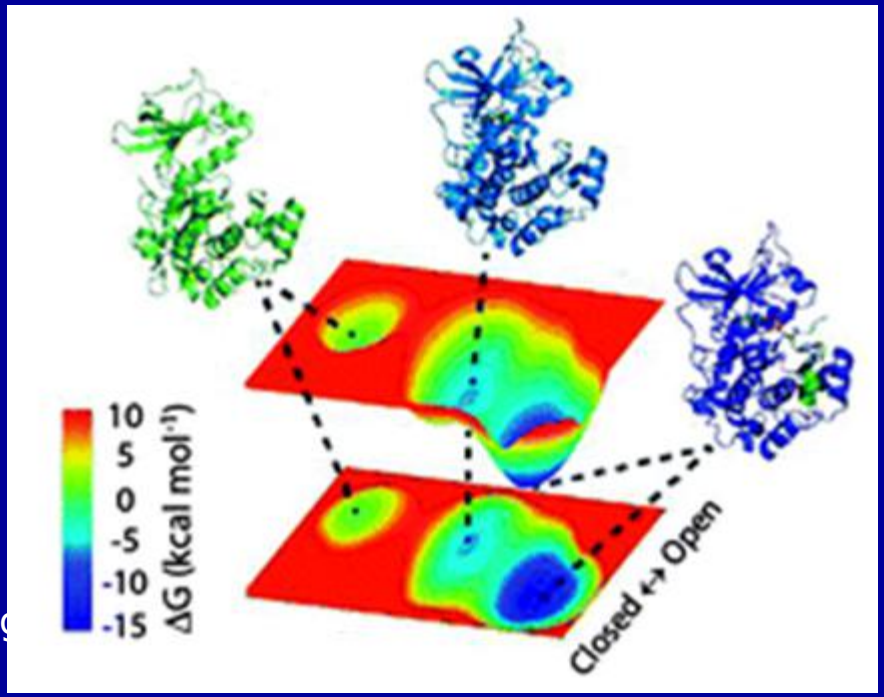
The energy of the system can be calculated using the forcefield. The conformation of the system can be altered to find lower energy conformations through a process called minimization.

Minimization algorithms:

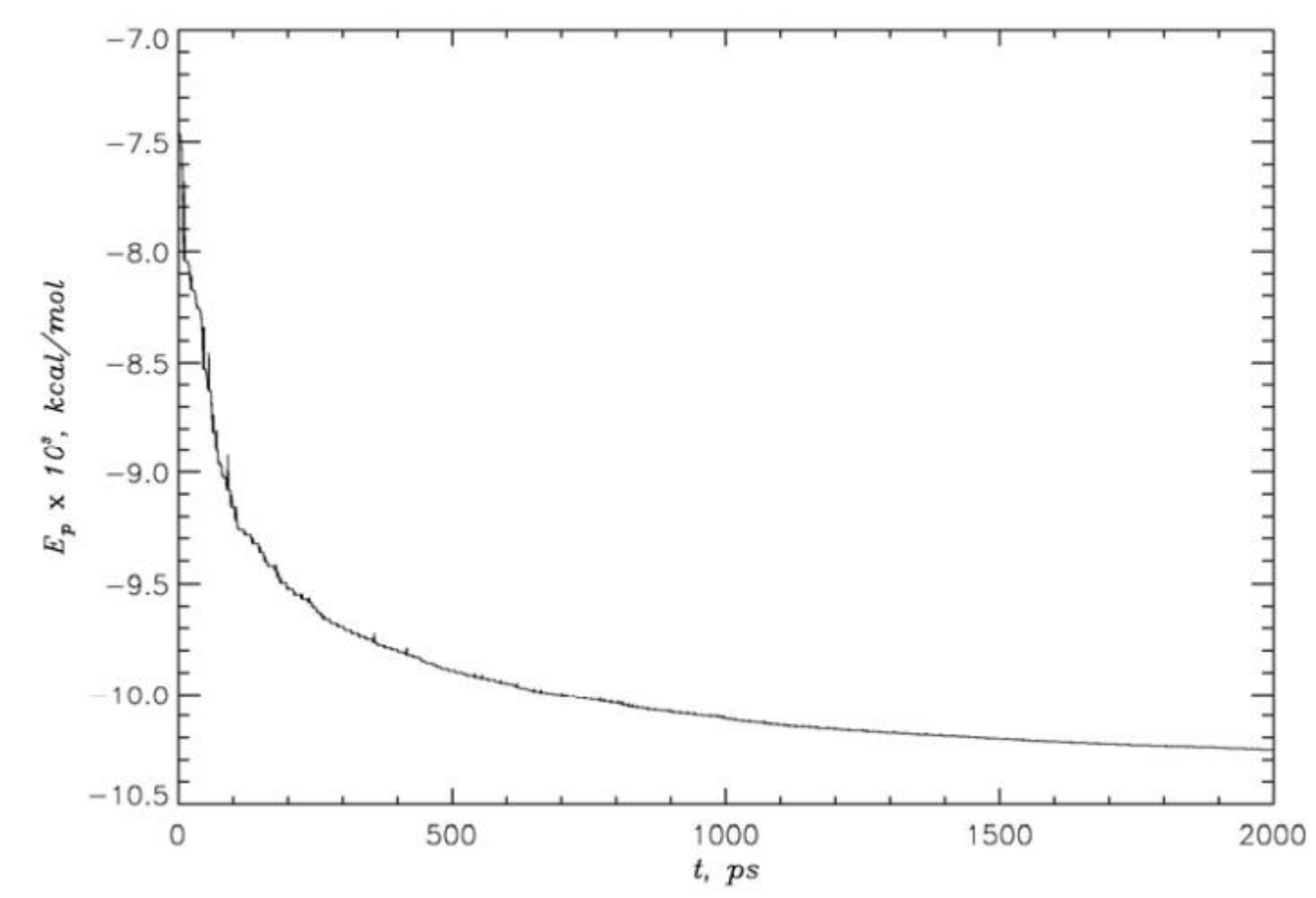
- steepest descent (slowly converging – use for highly restrained systems
- conjugate gradient (efficient, uses intelligent choices of search direction – use for large systems)
- BFGS (quasi-newton variable metric method)
- Newton-Raphson (calculates both slope of energy and

The above algorithms are deterministic, usually only lead to local minimum





Stochastic algorithms can also be used to continue the minimization.
But this is usually not practiced.



An example of energy minimization

7. Create the file **01_Min.in** that includes the following settings for

```
Minimize
 &cntrl
  imin=1,
  ntx=1,
  irest=0,
  maxcyc=2000,
  ncyc=1000,
  ntpr=100,
  ntwx=0,
  cut=8.0,
 /
```

The settings can be summarized as follows:

imin=1	Choose a minimization run
ntx=1	Read coordinates but not velocities from ASCII formatted **inpcrd** coordinate file
irest=0	Do not restart simulation. (not applicable to minimization)
maxcyc=2000	Maximum minimization cycles
ncyc=1000	The steepest descent algorithm for the first 0-**ncyc** cycles, then switches the conjugate gradient algorithm for **ncyc-maxcyc** cycles
ntpr=100	Print to the Amber **mdout** output file every **ntpr** cycles
ntwx=0	No Amber **mdcrd** trajectory file written (not applicable to minimization)
cut=8.0	Nonbonded cutoff distance in Angstroms

8. Run the minimization of alanine dipeptide with **sander**.

```
$ $AMBERHOME/bin/sander -O -i 01_Min.in -o 01_Min.out -p prmtop -c inpcrd -r 01_Min.rst \
-inf 01_Min.mdinfo
```

sander uses a consistent syntax for each step of MD simulation. Here is a summary of the command line options of **sander**:

-O	Overwrite the output files if they already exist
-i 01_Min.in	Choose input file (default **mdin**)
-o 01_Min.out	Write output file (default **mdout**)
-p prmtop	Choose parameter and topology file **prmtop**
-c inpcrd	Choose coordinate file **inpcrd**
-r 01_Min.rst	Write output restart file with coordinates and velocities (default **restrt**)
-inf 01_Min.mdinfo	Write MD info file with simulation status (default **mdinfo**)

sander should complete the minimization in a moderate amount of time (~ 27 seconds) depending on your computer specifications.

After sander completes, there should be an output file **01_Min.out**, a restart file **01_Min.rst**, and a MD info file **01_Min.mdinfo**. You will use the restart file **01_Min.rst** for the heating of the system.

Heating

Temperature

It measures kinetic energies of individual atoms.

It is actually averaged kinetic energy.

1. First we calculate the total kinetic energy

$$E_{\text{kin}} = \frac{1}{2}\sum_{i=1}^{N} m_i {v_i}^2$$

2. Then we divide N to get the average kinetic energy

$E_{kin}/_N$

3. It is almost the temperature T

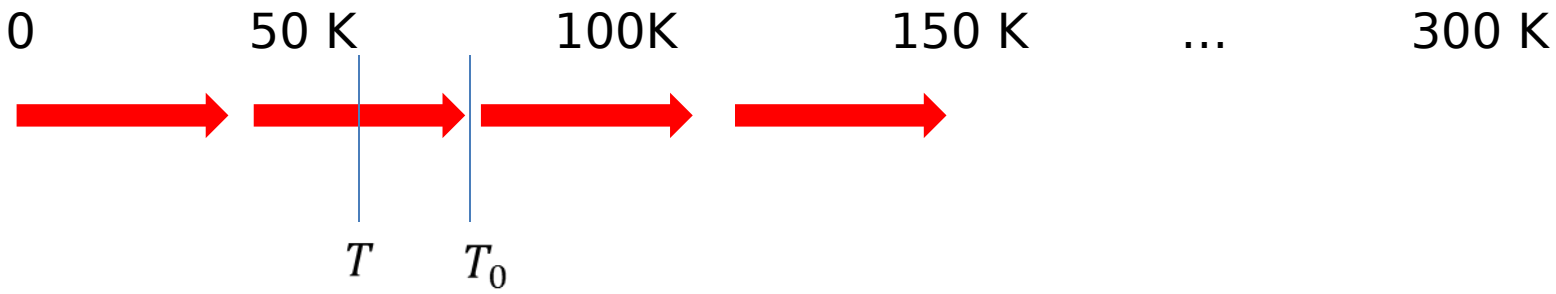
$$\frac{3}{2}k_B T = \frac{E_{kin}}{N} \qquad T = \frac{2E_{kin}}{3k_B N}$$

Heating

Let the temperature T gradually increase from 0 K to 300 K / 310 K

Initial crystal structure is static, which corresponds to 0K

Many target temperatures



How does the algorithm achieves the increase from T to T_0 ?

Basic Computation

The classical MD simulations boil down to numerically integrating Newton's equations of motion for the particles (atoms, in the simplest case) which build up the investigated system:

$$m\frac{d^2\mathbf{r}_i}{dt^2} = \mathbf{F}_i(\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_N), \quad i = 1, 2, \ldots, N.$$

Here $\mathbf{r}_i$ are the position vectors and $\mathbf{F}_i$ are the forces acting upon the N particles in the system.

Quite often forces derive from *potential functions*, $U(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$, representing the potential energy of the system for the specific geometric arrangement of the particles:

$$\mathbf{F}_i(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = -\nabla_{\mathbf{r}_i} U(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N).$$

This form implies the conservation of the total energy $E = E_{\text{kin}} + U$, where E_{kin} is the instantaneous kinetic energy.

Therefore we have

$$m\frac{d^2\mathbf{r}_i}{dt^2} = -\nabla_{\mathbf{r}_i} U(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \quad i = 1, 2, \ldots, N.$$

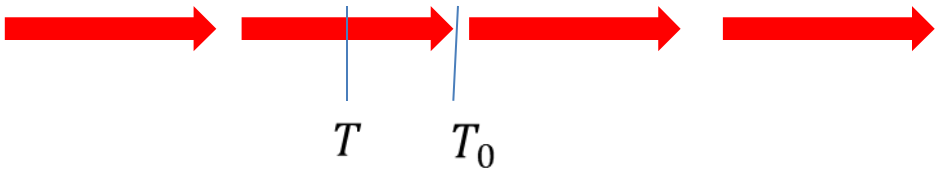
Because it is temperature to be discussed, we use velocity for further discu

$$\boldsymbol{v}_i = \frac{d\boldsymbol{r}_i}{dt}$$

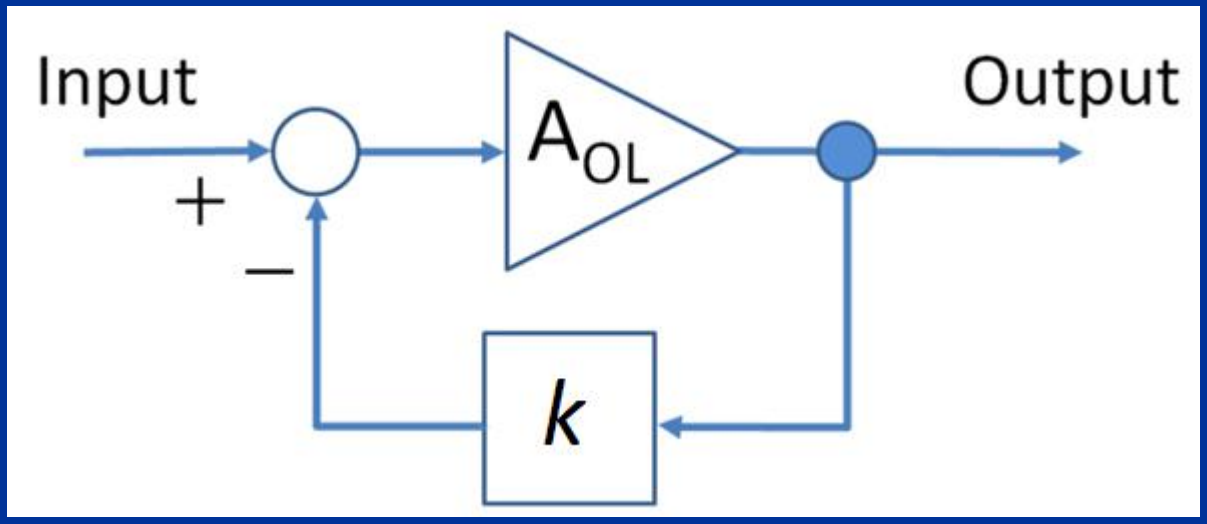
$$m\frac{d\boldsymbol{v}_i}{dt} = -\nabla_{\boldsymbol{r}_i} U(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N)$$

Once initial positions and velocities given, the evolution of them are deterr

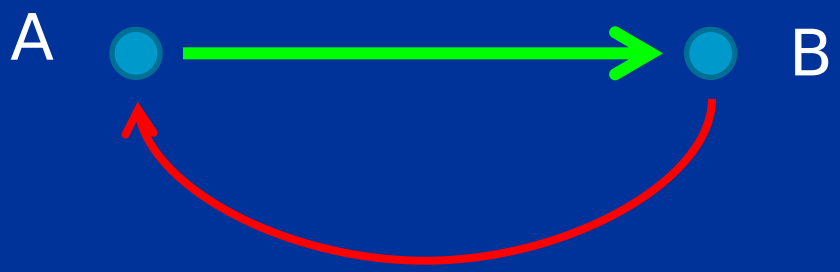
But we have to control it --- Increase the velocity so that T can rise to T_0



Negative feedback (to achieve self-limiting)

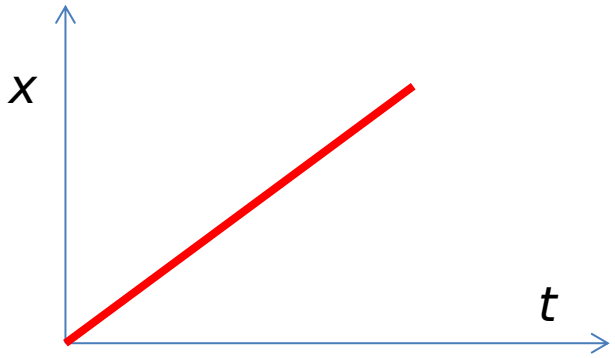


The most fundamental regulatory mechanism
Widely used by nature and in the engineering syste



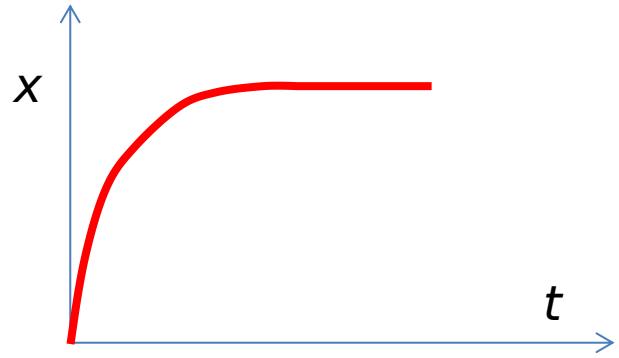
Without feedback

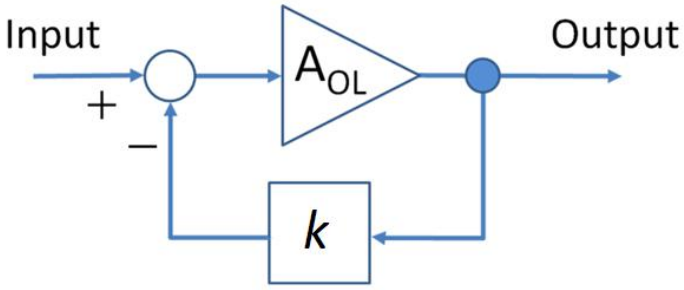
$$\frac{dx}{dt} = 1$$

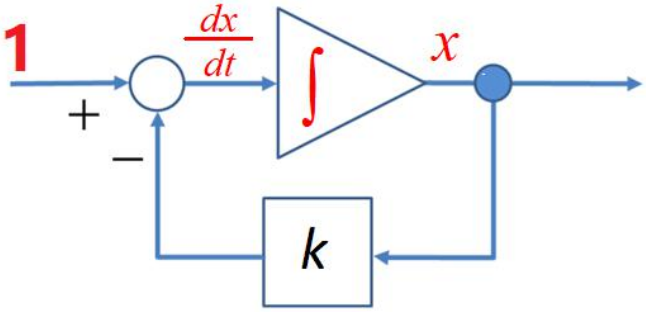


With feedback

$$\frac{dx}{dt} = 1 - kx$$







$$m\frac{d\boldsymbol{v}_i}{dt} = -\nabla_{\boldsymbol{r}_i} U(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N) \; -\gamma M \boldsymbol{v}_i$$

$$\gamma = \frac{1}{2\tau}\left(1 - \frac{T_0}{T}\right)$$

$$T = \frac{2E_{kin}}{3k_B N} \quad = \frac{1}{3k_B N}\sum_{i=1}^{N} m_i {v_i}^2$$

9. Create the file **02_Heat.in** that includes the following settings for heating:

```
Heat
 &cntrl
  imin=0,
  ntx=1,
  irest=0,
  nstlim=10000,
  dt=0.002,
  ntf=2,
  ntc=2,
  tempi=0.0,
  temp0=300.0,
  ntpr=100,
  ntwx=100,
  cut=8.0,
  ntb=1,
  ntp=0,
  ntt=3,
  gamma_ln=2.0,
  nmropt=1,
  ig=-1,
 /
&wt type='TEMP0', istep1=0, istep2=9000, value1=0.0, value2=300.0 /
&wt type='TEMP0', istep1=9001, istep2=10000, value1=300.0, value2=300.0 /
&wt type='END' /
```

The settings can be summarized as follows:

imin=0	Choose a molecular dynamics (MD) run [no minimization]
nstlim=10000	Number of MD steps in run (nstlim * dt = run length in ps)
dt=0.002	Time step in picoseconds (ps). The time length of each MD step
ntf=2	Setting to not calculate force for SHAKE constrained bonds
ntc=2	Enable SHAKE to constrain all bonds involving hydrogen
tempi=0.0	Initial thermostat temperature in K (see **NMROPT** section)
temp0=300.0	Final thermostat temperature in K (see **NMROPT** section)
ntwx=1000	Write Amber trajectory file **mdcrd** every **ntwx** steps
ntb=1	Periodic boundaries for constant volume
ntp=0	No pressure control
ntt=3	Temperature control with Langevin thermostat
gamma_ln=2.0	Langevin thermostat collision frequency
nmropt=1	NMR restraints and weight changes read (see **NMROPT** section)
ig=-1	Randomize the seed for the pseudo-random number generator [always a good idea unless you are debugging a simulation problem]

10. Run the heating of alanine dipeptide with **sander**

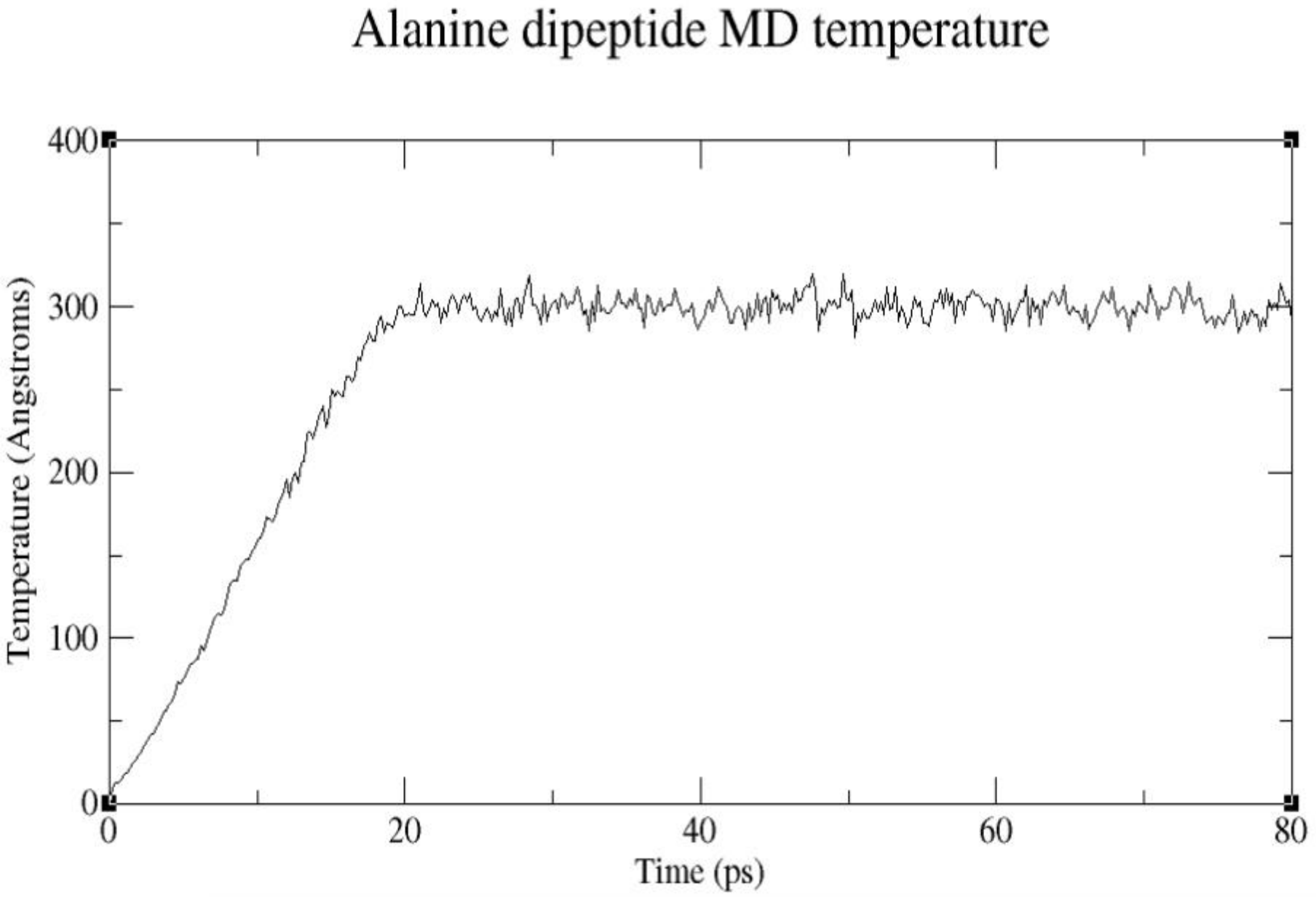
```
$ $AMBERHOME/bin/sander -O -i 01_Min.in -o 01_Min.out -p prmtop -c inpcrd -r 01_Min.rst \
-inf 01_Min.mdinfo
```

```
$ $AMBERHOME/bin/sander -O -i 02_Heat.in -o 02_Heat.out -p prmtop -c 01_Min.rst \
-r 02_Heat.rst -x 02_Heat.mdcrd -inf 02_Heat.mdinfo
```

Here is a summary of the command line options for **sander:**

-c 01_Min.rst	Now for the input coordinates we choose the restart file from minimization
-x 02_Heat.mdcrd	Output trajectory file for MD simulation (default **mdcrd**)

sander should complete the heating in a moderate amount of time (~ 2.5 mins) depending on your computer specifications.


Alanine dipeptide MD temperature
Temperature (Angstroms)
400
300
200
100
0
0
20
40
60
80
Time (ps)

Run production MD

11. Now that minimization and heating are complete.
    We move on to the actual production MD.

```
$ $AMBERHOME/bin/sander -O -i 03_Prod.in -o 03_Prod.out -p prmtop -c 02_Heat.rst \
-r 03_Prod.rst -x 03_Prod.mdcrd -inf 03_Prod.info &
```

Note: With the "&" at the end of the command, **sander** now runs in the background

Now **sander** is running in the background. It will take some time to run the production MD.

12. Create the file **03_Prod.in** with the settings for production MD

```
Production
 &cntrl
  imin=0,
  ntx=5,
  irest=1,
  nstlim=30000,
  dt=0.002,
  ntf=2,
  ntc=2,
  temp0=300.0,
  ntpr=100,
  ntwx=100,
  cut=8.0,
  ntb=2,
  ntp=1,
  ntt=3,
  gamma_ln=2.0,
  ig=-1,
 /
```

The settings for production can be summarized as follows:

irest=1	Restart previous MD run [This means velocities are expected in the inpcrd file and will be used to provide initial atom velocities]
temp0=300.0	Thermostat temperature. Run at 300K
ntb=2	Use periodic boundary conditions with constant pressure
ntp=1	Use the Berendsen barostat for constant pressure simulation
ntx=5	Read coordinates and velocities

from which file?

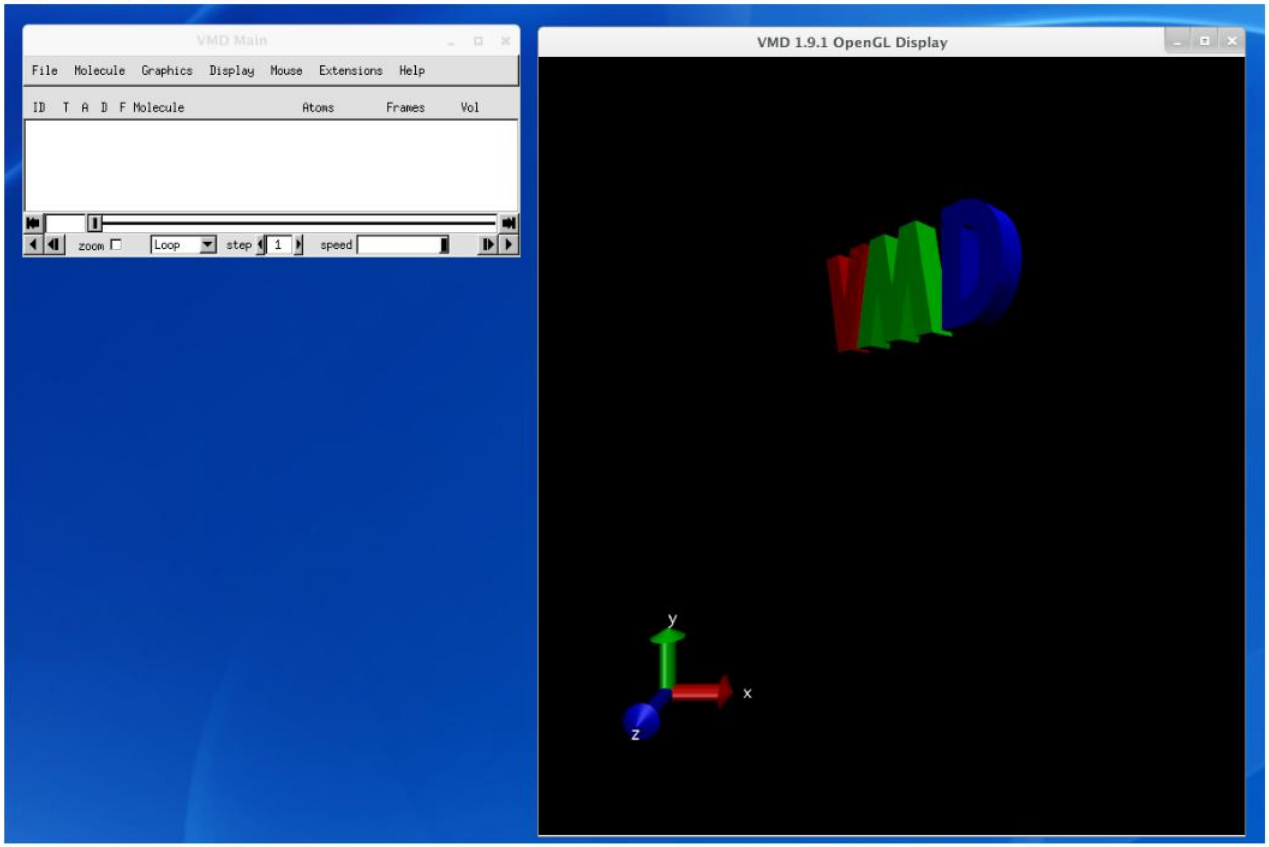
```
$ $AMBERHOME/bin/sander -O -i 03_Prod.in -o 03_Prod.out -p prmtop -c 02_Heat.rst \
-r 03_Prod.rst -x 03_Prod.mdcrd -inf 03_Prod.info &
```

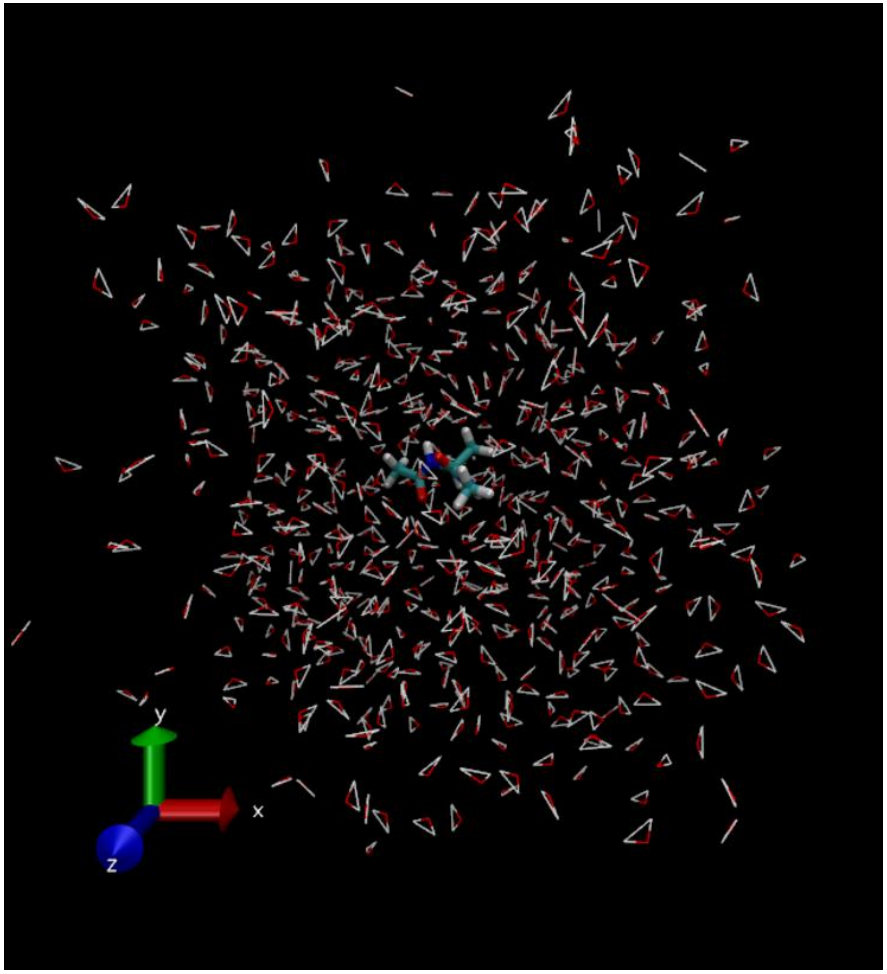
Visualize the results

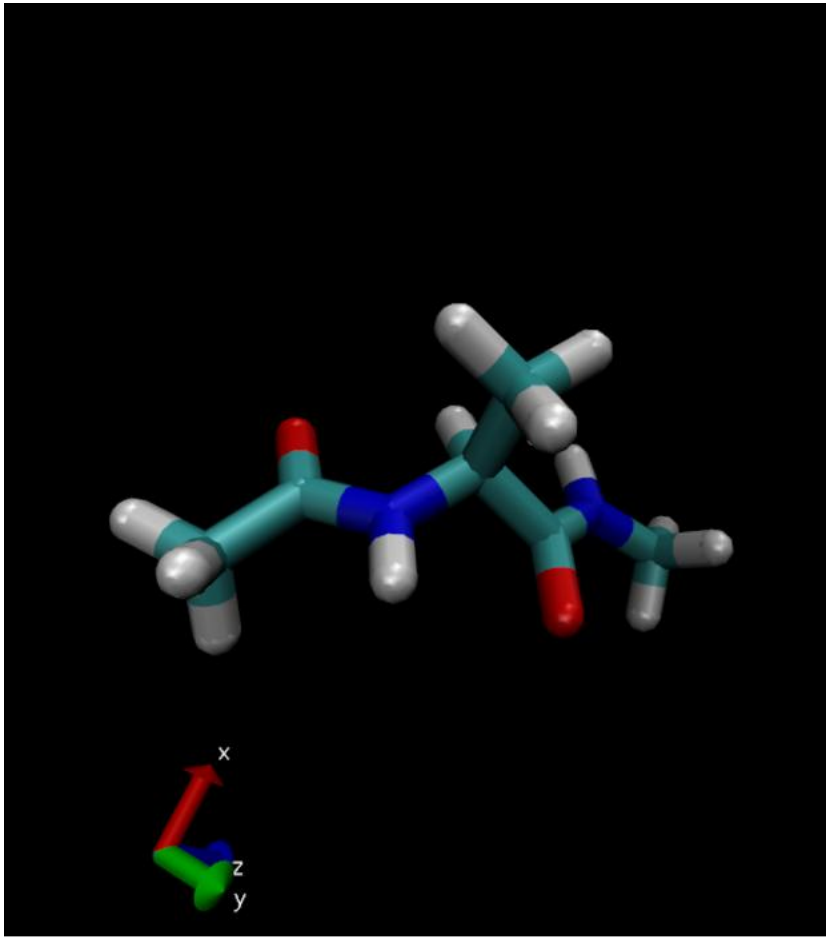
You've now run an MD simulation. In order to visualize the results, we will now use a program called VMD (Visual Molecular Dynamics). This is a molecular graphics program that can render molecular structures in 3D. VMD not only loads Protein Database (PDB) structure files, but also MD trajectories from many programs

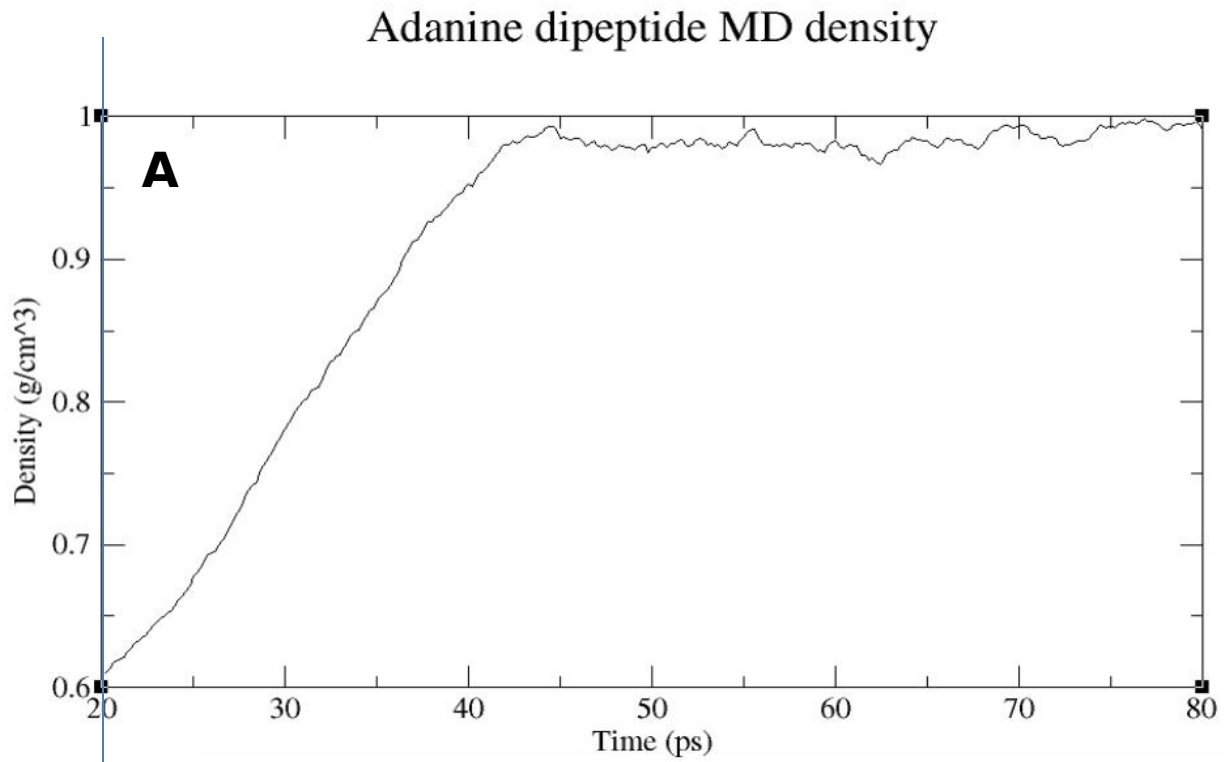
```
$ vmd
```

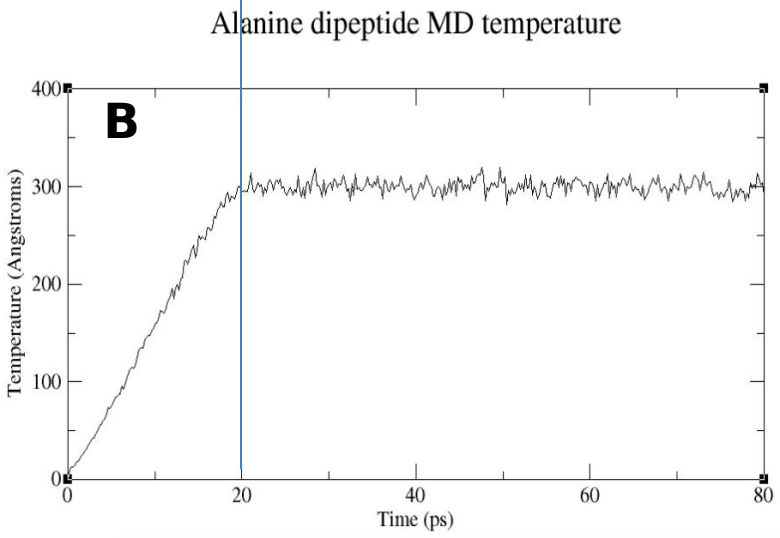
VMD should look like this:




y
x
z


x
z
y



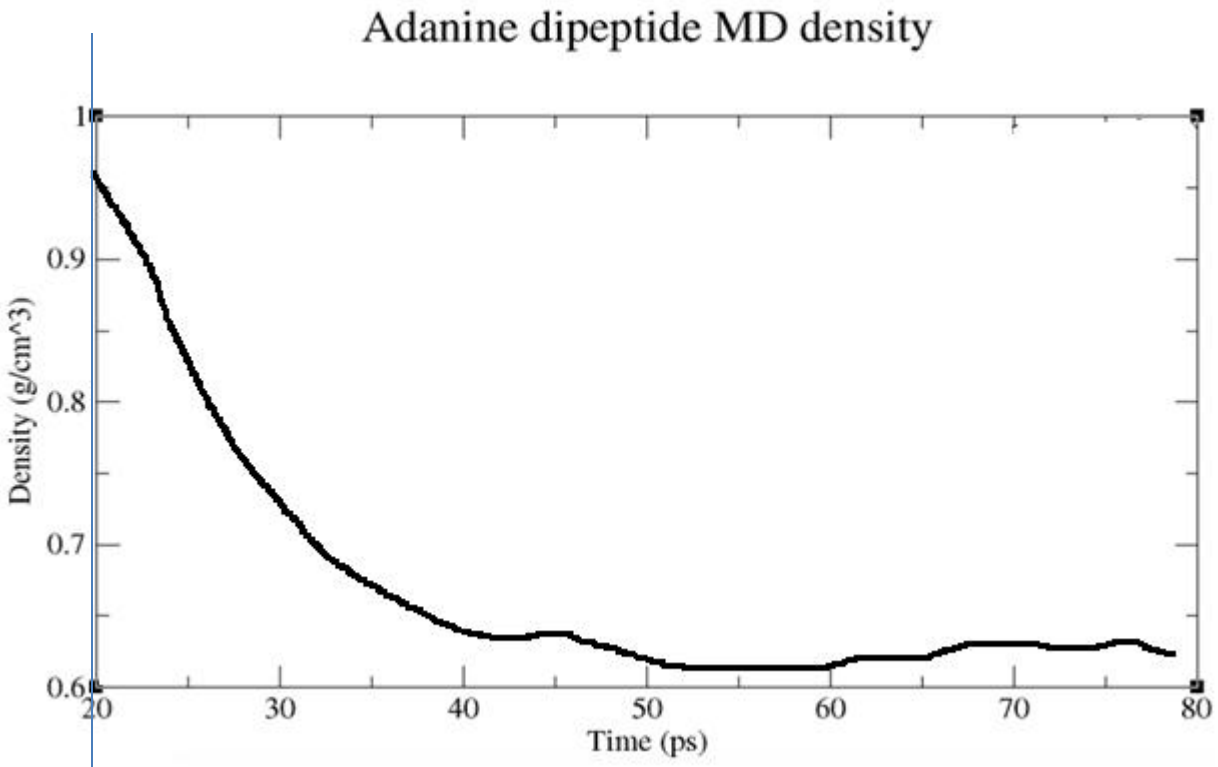


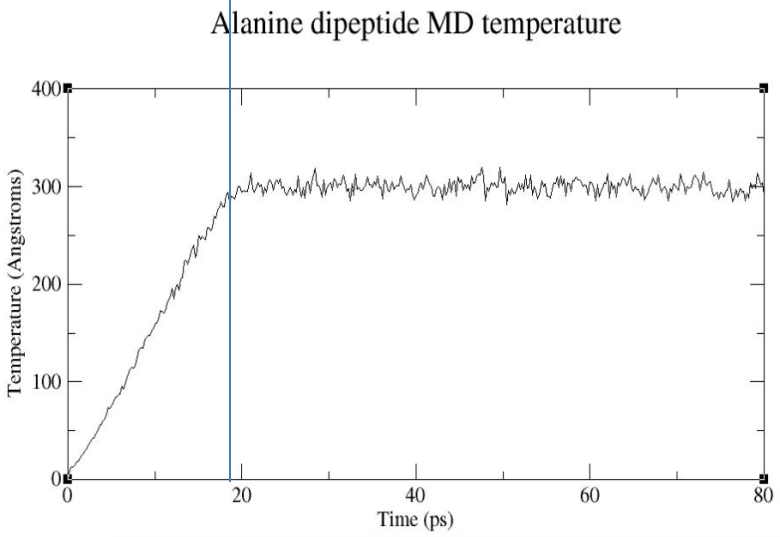
What happened? Why density increases?

This implies the volume of Box reduces!

This in turn implies that after heating,
The pressure is still smaller than that set for the production phase.

So the box shrinks so that the pressure inc
to the preset one.



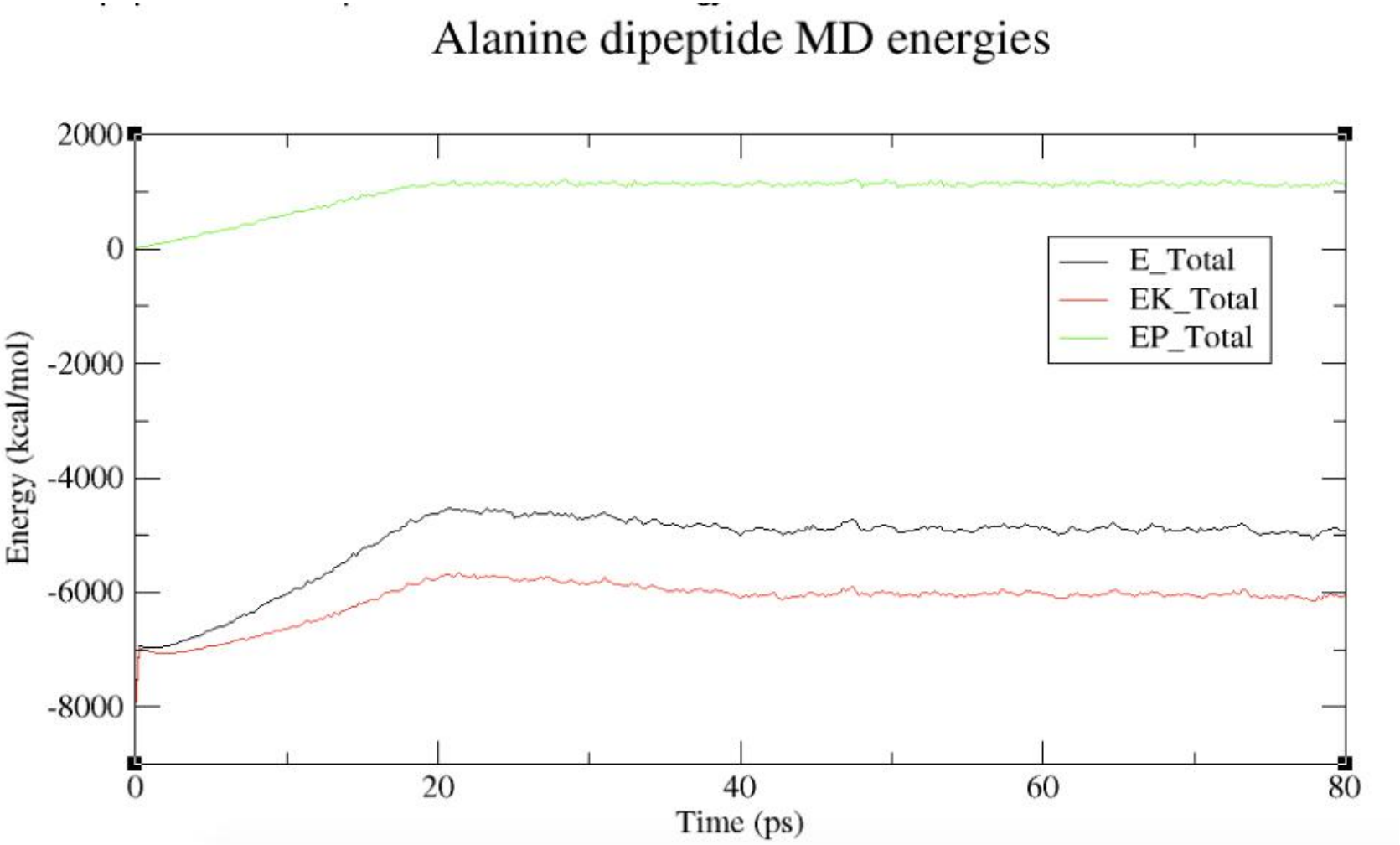


Can this happen? Density decreases?

This implies the volume of Box increases

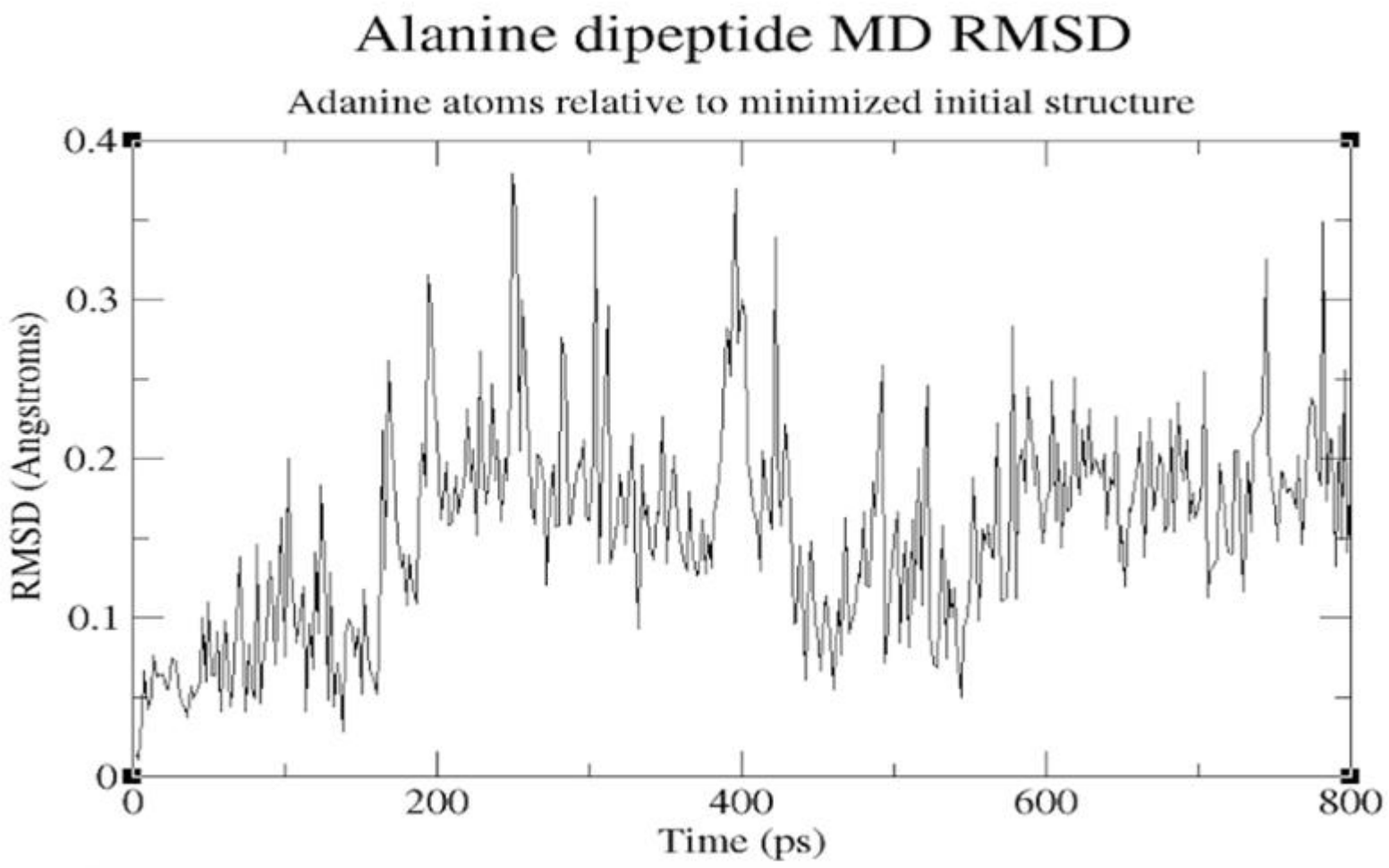
This in turn implies that after heating, The pressure is larger than that set for the production phase.

So the box expands so that the pressure decreases to the preset one.



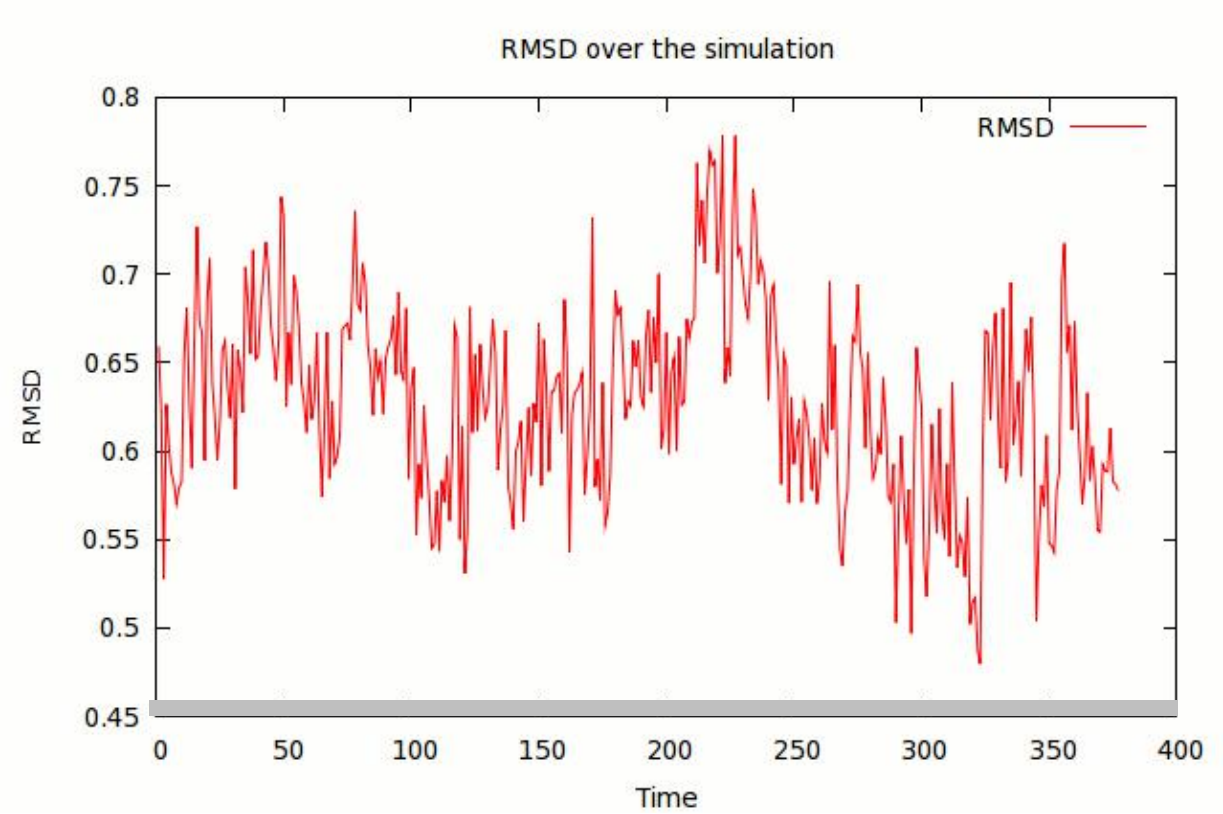
This plot shows the total system energy which can be decomposed to the total potential energy and the total kinetic energy.

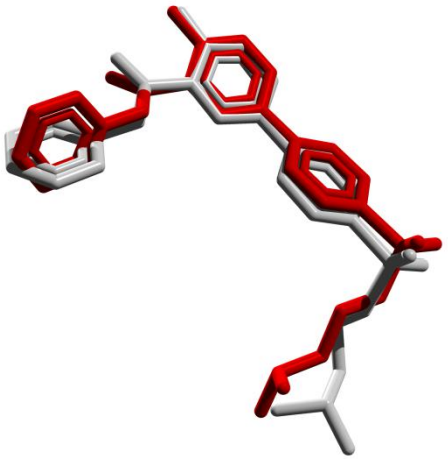
Alanine dipeptide MD RMSD relative to minimized initial structure



RMSD: Root-mean-square deviation of atomic positions

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2}$$





Since the next example is a larger protein, we introduce

1. How to get protein structure
2. PDB files.