Hands-on training of MD simulation

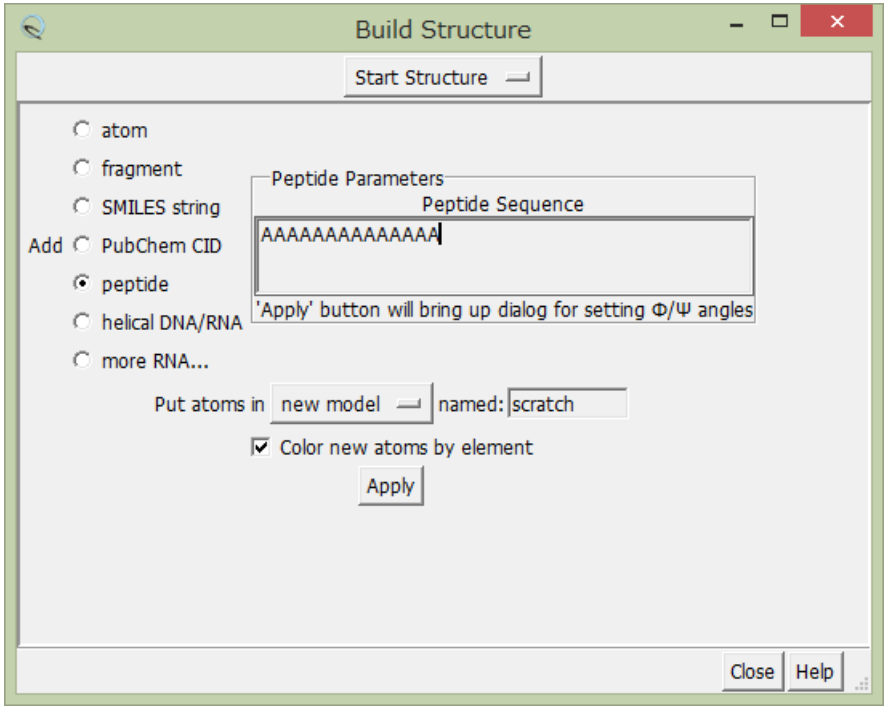
Agricultural Bioinformatics Research Unit,
Graduate School of Agricultural and Life Sciences, The University of Tokyo
Tohru Terada

Today's topics

- MD simulation of a peptide
- Modeling of the aqueous environment
- MD simulation of a peptide
  - Assignment 1
- MD simulation of a protein
  - Assignment 2
- Acceleration of a simulation
- Note

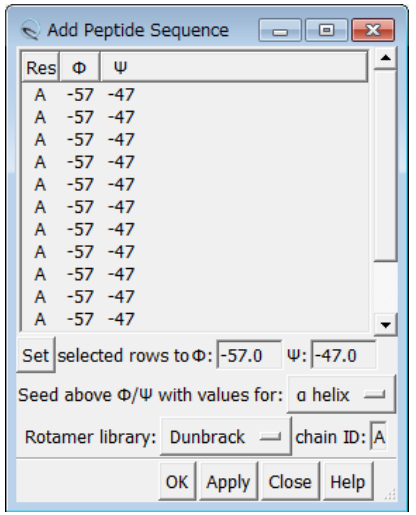
Generating a peptide (1)

1. Start UCSF Chimera 1.10.1
2. Choose “Tools” → “Structure Editing” → “Build Structure.”
3. Choose “Start Structure” and “peptide,” put 14 A’s in Peptide Sequence, and click “Apply.”



Generating a peptide (2)

4. In the Add Peptide Sequence window, you can change backbone dihedral angles. Here, simply click “OK.”
5. A peptide structure appears in the main window. Choose “Actions” → “Atoms/Bonds” → “show” and “Actions” → “Ribbon” → “hide.”
6. Choose “File” → “Save PDB” and save the structure as “ala14.pdb” in the Desktop.



Assignment of force-field parameters

1. Choose “Tools” → “Structure Editing” → “AddH” and click “OK” to add hydrogens.
2. Choose “Tools” → “Structure Editing” → “Add Charge,” change the force field of the Standard residues to “AMBER ff99SB” and click “OK.”
3. Choose “Tools” → “Amber” → “Write Prmtop” to save the parameter/topology file. Change the Folder to “C:¥Users¥iu¥Desktop”, set the File name to “ala14,” change the force field type to “AMBER ff99SB,” and click “Save.”

Procedure of MD simulation

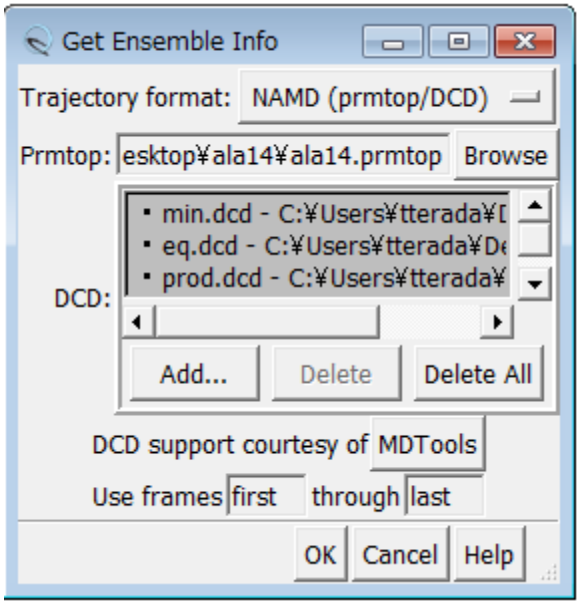
1. Download namd2.exe and tcl85t.dll from the web page of this lecture and save them in the desktop.
2. Download ala14.zip from the web page of this lecture and save it in the desktop.
3. Unzip ala14.zip by double-clicking it. Move ala14.prmtop and ala14.inpcrd to the generated folder, ala14.
4. Run run.bat by double-clicking it.

Software

- NAMD
  - Used here. Free.
  - AMBER and CHARMM force fields are supported.
  - http://www.ks.uiuc.edu/Research/namd/
- Gromacs
  - Free.
  - AMBER, CHARMM, GROMOS force fields are supported.
  - http://www.gromacs.org/
- AMBER, CHARMM, etc.

Visualizing the result (1)

1. Start Chimera. (If already running, choose “File” → “Close Session.”)
2. Choose “Tools” → “MD/Ensemble Analysis” → “MD Movie.”
3. Change Trajectory format to “NAMD (prmtop/DCD),” set Prmtop to ala14.prmtop, add min.dcd, eq.dcd, prod.dcd in this order to the list of the DCD files, and click “OK.”



Visualizing the result (2)

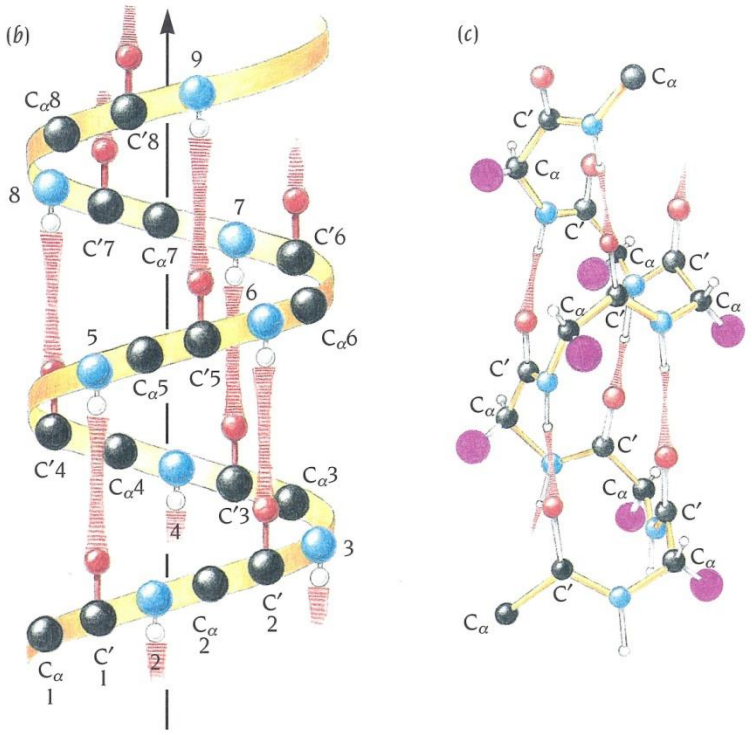
4. Click the play button in the MD Movie window to watch the movie.
5. Choose “Actions” → “Atoms/Bonds” → “show” and “Actions” → “Ribbon” → “hide” to change the representation to the stick model.
6. To correct the error in the atom coloring, choose “File” → “Open” and open color.com. Watch the movement of the atoms.

Calculating RMSD

1. Choose “Select” → “Atom Specifier,” put “@CA” in the textbox of Atom specifier to select, and click “OK” to select Cα atoms.
2. Choose “Analysis” → “Plot” → “RMSD” from the menu of the MD Movie window.
3. Change Ignore hydrogens to “false” and click “Plot.”

Measuring hydrogen-bond length

1. Choose “Analysis” → “Plot” → “Distances” in the MD Movie window.
2. Click on one atom of the pair of interest and click on the other atom while pressing the Ctrl key.
3. Click the “Plot” button in the MD Plots window.

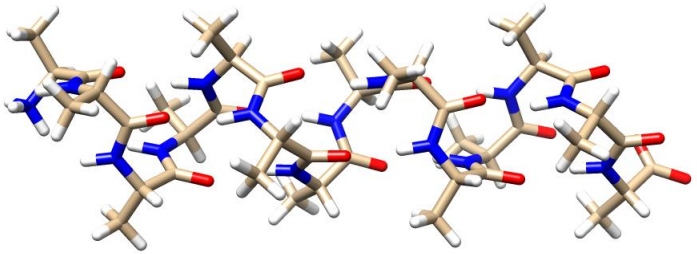


In α helices, hydrogen bonds are formed between carbonyl oxygen of the *i*-th residue and amide nitrogen of the (*i*+4)-th residue.

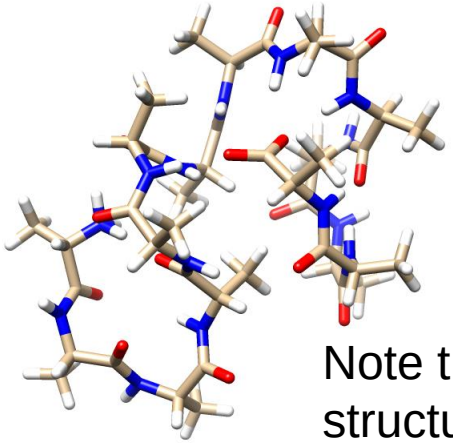
Branden & Tooze “Introduction to Protein Structure” 2nd Ed.

Result of the simulation

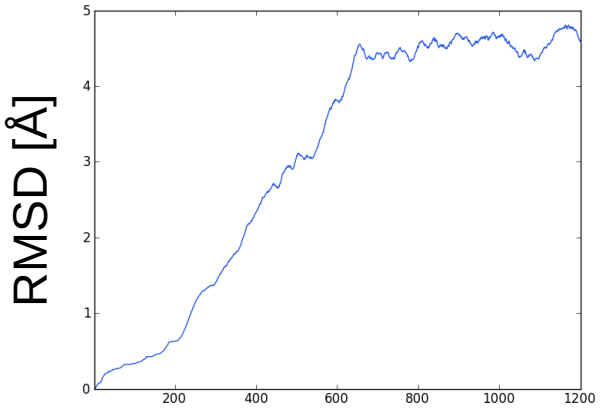
Initial structure

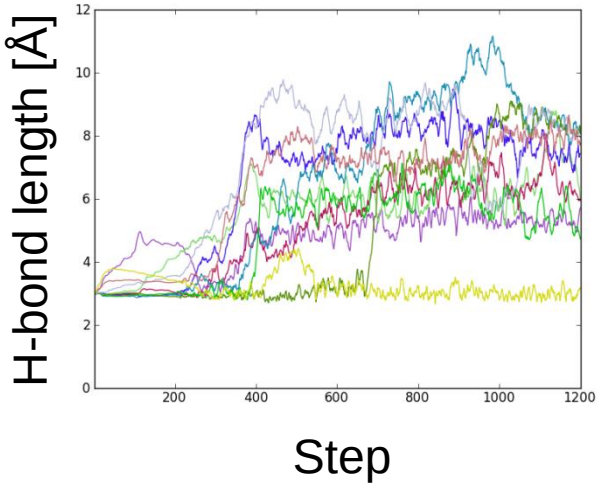


Final structure



Note that the α helix structure is broken by the simulation.





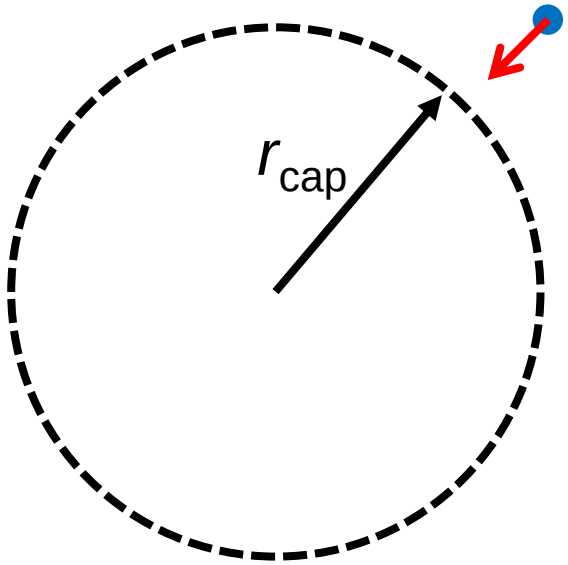
Modeling the aqueous environment (1)

- The simulation was performed in vacuum, and therefore the solvation effect by water molecules is not considered.
- When performing biomacromolecular simulations, it is necessary to put the molecules in an aqueous environment.

Modeling the aqueous environment (2)

- Explicit water model
  - Place water molecules in a sphere
  - Place water molecules in a rectangular box →used under periodical boundary condition
- Implicit water model to approximate solvation free energy
  - Non-polar term→proportional to solvent accessible surface area
  - Polar term→continuum dielectric model
    - Poisson-Boltzmann equation
    - Generalized Born model

Spherical placement



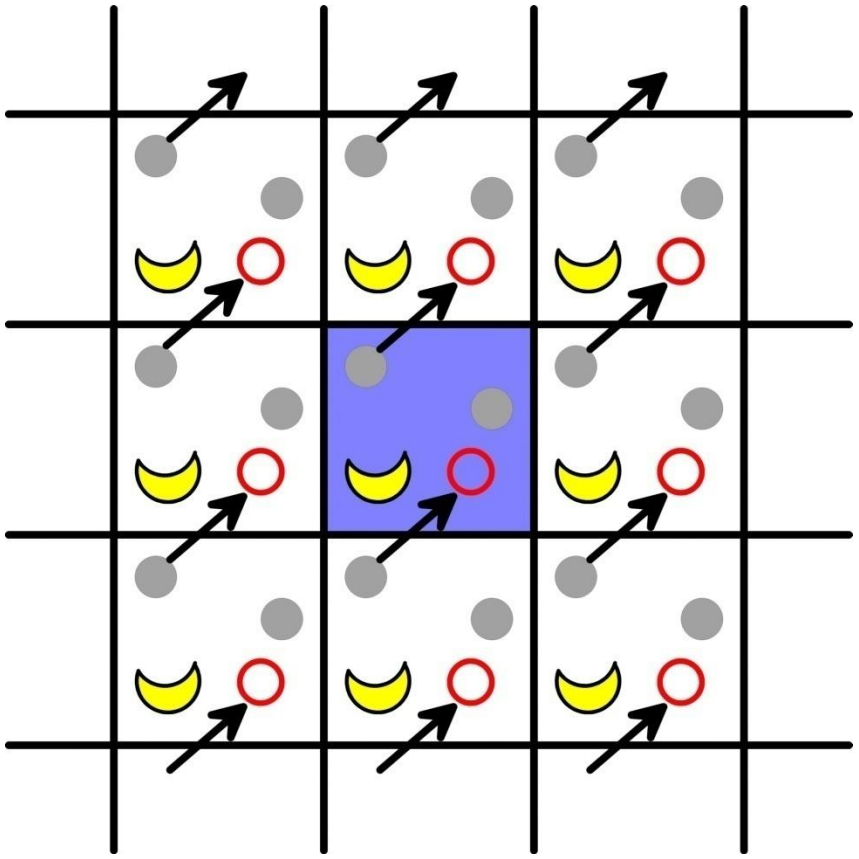
$$r = \sqrt{\left(x - x_0\right)^2 + \left(y - y_0\right)^2 + \left(z - z_0\right)^2}$$

$$E_{cap} = \begin{cases} 0 & r < r_{cap} \\ k\left(r - r_{cap}\right)^2 & r \geq r_{cap} \end{cases}$$

- To prevent vaporization of the water molecules, a restraining force is applied to the water molecule going outside of the sphere of radius r_{cap}.
- Surroundings of the water molecules near the surface of the sphere are different from those of the water molecules near the center of the sphere.

Periodical boundary condition

- The central cell is replicated to form an infinite lattice.
- A molecule that goes out the cell enters through the opposite face.
- All the molecules feel the same surroundings.
- The dimensions of the cell should be large enough to reduce the unnatural effect from neighboring cells.



Calculation of pressure

$$F = E - TS, \quad dE = -PdV + TdS$$

$$dF = dE - TdS - SdT = -PdV - SdT$$

$$P = -\left(\frac{\partial F}{\partial V}\right)_T = k_B T\left(\frac{\partial \ln Z}{\partial V}\right)_T = \frac{k_B T}{Z}\left(\frac{\partial Z}{\partial V}\right)_T$$

$$= \frac{Nk_B T}{V} + \frac{1}{3V}\sum_{i=1}^{N} \mathbf{r}_i \cdot \mathbf{f}_i$$

virial theorem

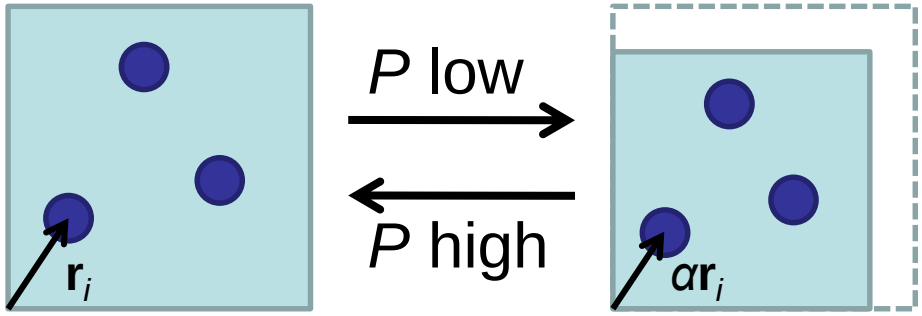
$$= \frac{Nk_B T}{V} + \frac{1}{3V}\sum_{i=1}^{N}\sum_{j=i+1}^{N} \mathbf{r}_{ij} \cdot \mathbf{f}_{ij}$$

Use this form under periodic boundary condition

In a system of non-interacting particles (ideal gas),
$PV = Nk_B T = nRT$

Pressure regulation

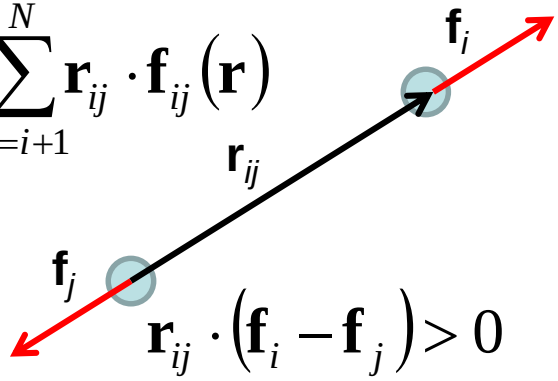
- Pressure is regulated by changing the dimensions of the cell under the periodical boundary condition.

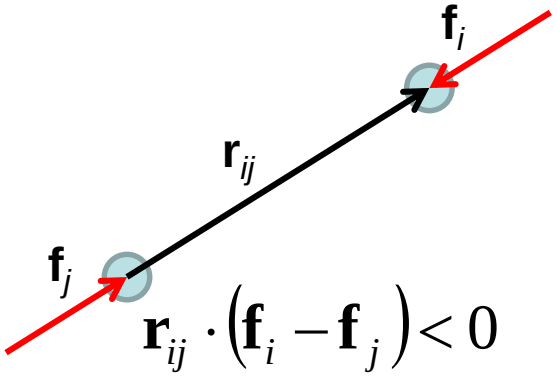


Center-of-mass positions of the molecules are scaled. The structure of each molecule does not change.

- Pressure can be negative.

$$P = \frac{Nk_{\mathrm{B}}T}{V} + \frac{1}{3V}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\mathbf{r}_{ij}\cdot\mathbf{f}_{ij}(\mathbf{r})$$





Simulation in aqueous environment (1)

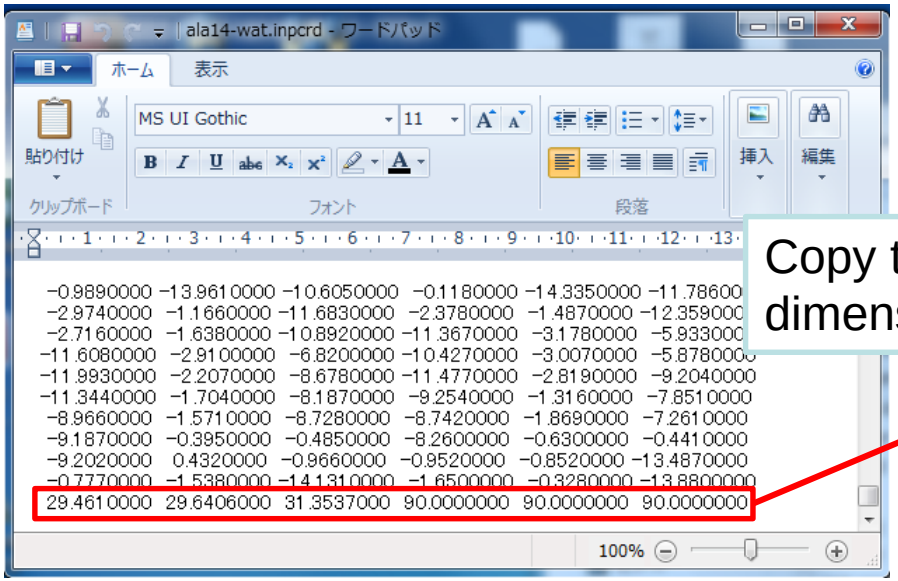
1. Start Chimera and open ala14.pdb.
2. Change the representation to sticks and add hydrogen atoms.
3. Choose “Tools” → “Structure Editing” → “Solvate.” Change Solvate method to “Box” and Solvent Model to “TIP3PBOX,” set Box size to “6,” and click “OK.”
4. Choose “Tools” → “Structure Editing” → “Add Charge.” Change the force field of Standard residues to “AMBER ff99SB,” and click “OK.”

Simulation in aqueous environment (2)

5. Choose “Tools” → “Amber” → “Write Prmtop.” Ensure the Folder is “C:¥Users¥iu¥Desktop,” set File name to “ala14-wat,” change the force field type to “AMBER ff99SB,” and click “Save.”
6. Choose “File” → “Save PDB” to save the solvated structure as ala14-wat.pdb in the desktop.
7. Download ala14-wat.zip from the web page of this lecture, save it in the desktop, and unzip it.
8. Move ala14-wat.prmtop, ala14-wat.inpcrd, and ala14-wat.pdb to the ala14-wat folder.

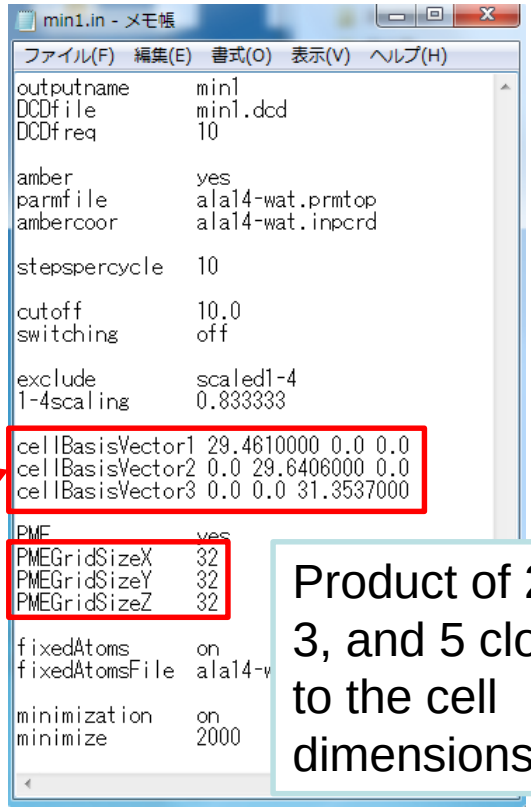
Simulation in aqueous environment (3)

9. Open the ala14-wat folder and run restraint.pl by double-clicking it to generate ala14-wat_rest.pdb.
10. Modify min1.in as follows:



Copy the cell dimensions.

Open ala14-wat.inpcrd in WordPad and go to the last line.



Product of 2, 3, and 5 close to the cell dimensions

Simulation in aqueous environment (4)

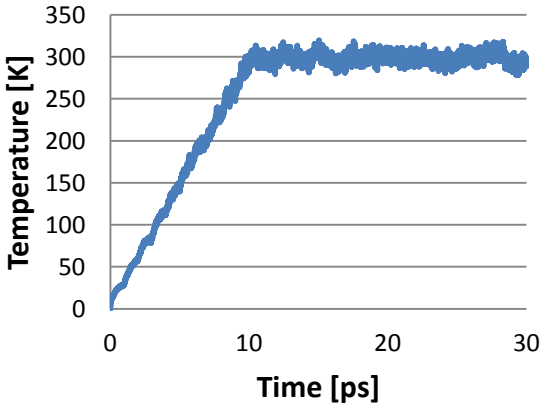
11. Run run.bat by double-clicking it to perform simulations in the following order (They will take about 3 minutes in total):
    ① Energy minimization (water only) (min1)
    ② Energy minimization (whole) (min2)
    ③ Equilibration (0→300 K) (eq1, 10 ps)
    ④ Equilibration (constant-*NPT*) (eq2, 10 ps)
    ⑤ Production (prod, 10 ps)

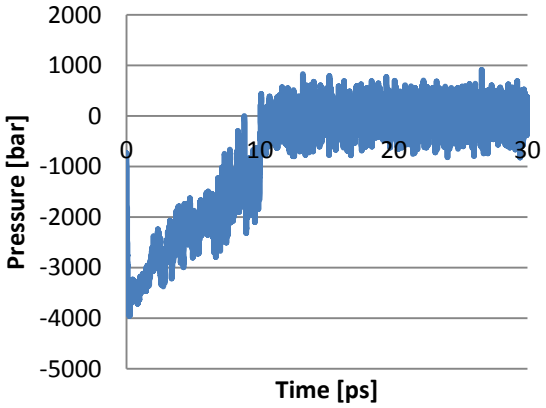
Analysis of the result

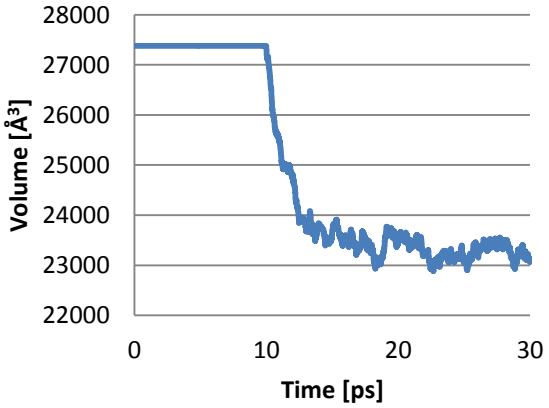
1. Start Command Prompt (cmd) and execute the following commands:

```
> cd Desktop¥ala14-wat
> energy.pl eq1.log eq2.log prod.log
```

2. Open energy.csv in Excel.

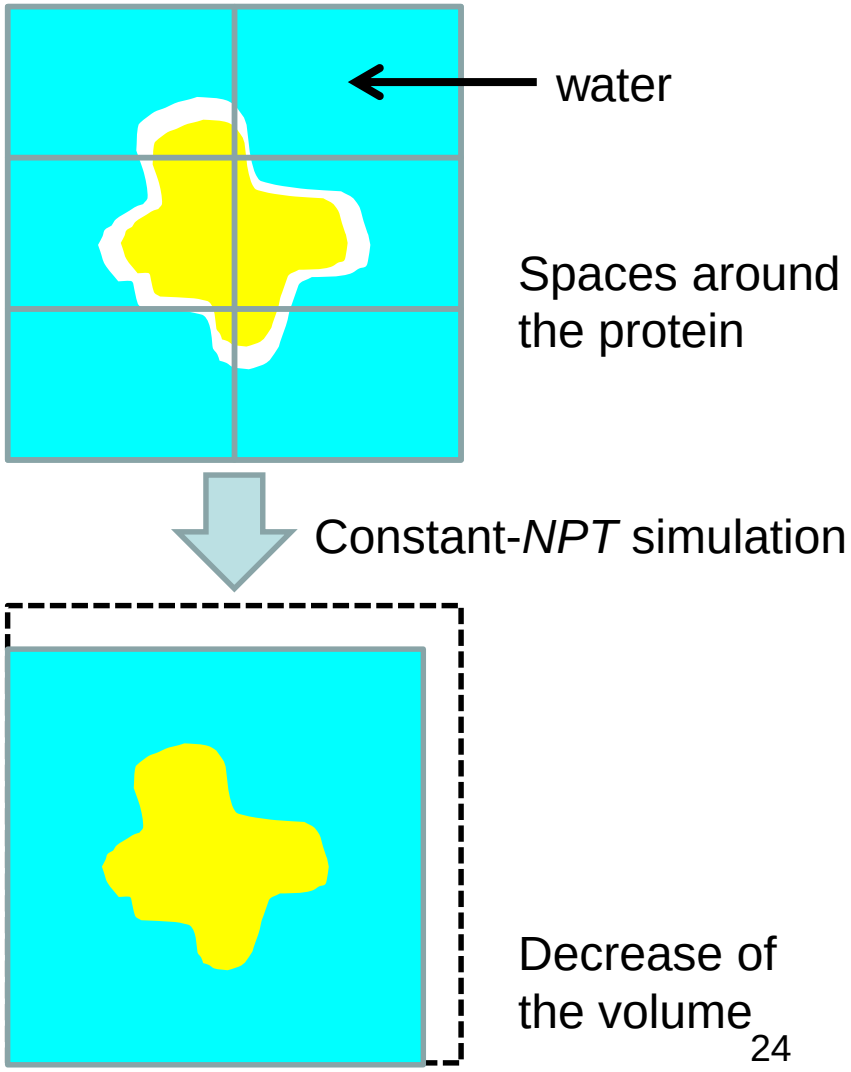






Reason for the volume change

- When water molecules are placed around the protein, a block of equilibrated water is used to fill the cell and the water molecules overlapping with protein atoms are automatically removed. As a result, spaces are produced around the protein.
- During the constant-*NPT* simulation, the water structure is optimized. As a result, the space is filled and the volume is decreased.



Assignment 1

- Plot the temperature (TEMP), pressure (PRESSURE), and volume (VOLUME) as a function of time during the equilibration runs (eq1 and eq2) and the production run (prod).
  - The time step Δt is 2 fs.
- Plot the variation of the Cα RMSD during the energy minimization (min1 and min2), equilibration runs (eq1 and eq2) and the production run (prod).
  - Set Ignore hydrogens to “false.”
- What conclusion(s) can you draw from these plots?

Computational time (1)

- Target: water spheres (the TIP3P model)
- The Sander module of Amber 11 is used.
- 8 cores of an Intel Xeon Processor are used.
- The time step Δt is 0.5 fs.
- Computational time of a 1-ps simulation is measured.

#atoms	T_{total} [s]	Ratio	T_{nb} [s]	T_{nb}/T_{total}
3087	35	1.0	35	0.983
6066	137	3.9	136	0.995
10608	420*	12.0	419	0.998

*4.9 days per ns

Acceleration of simulation

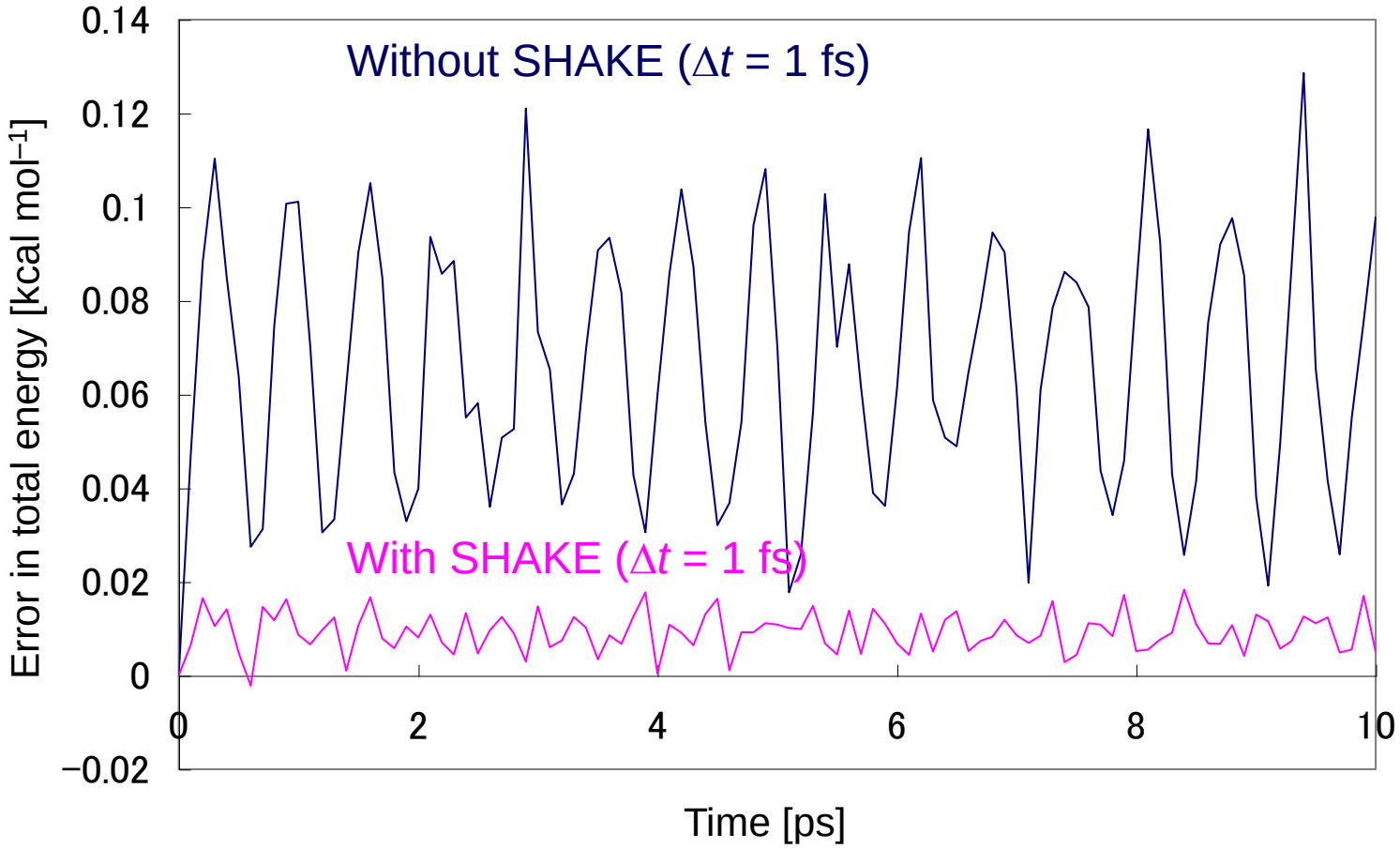
- Use of a larger Δt
  - The SHAKE method
  - The multiple time step metod
- Approximation of non-bonded interactions
  - The cut-off method
  - The fast multipole method
  - The particle mesh Ewald (PME) method

The SHAKE method

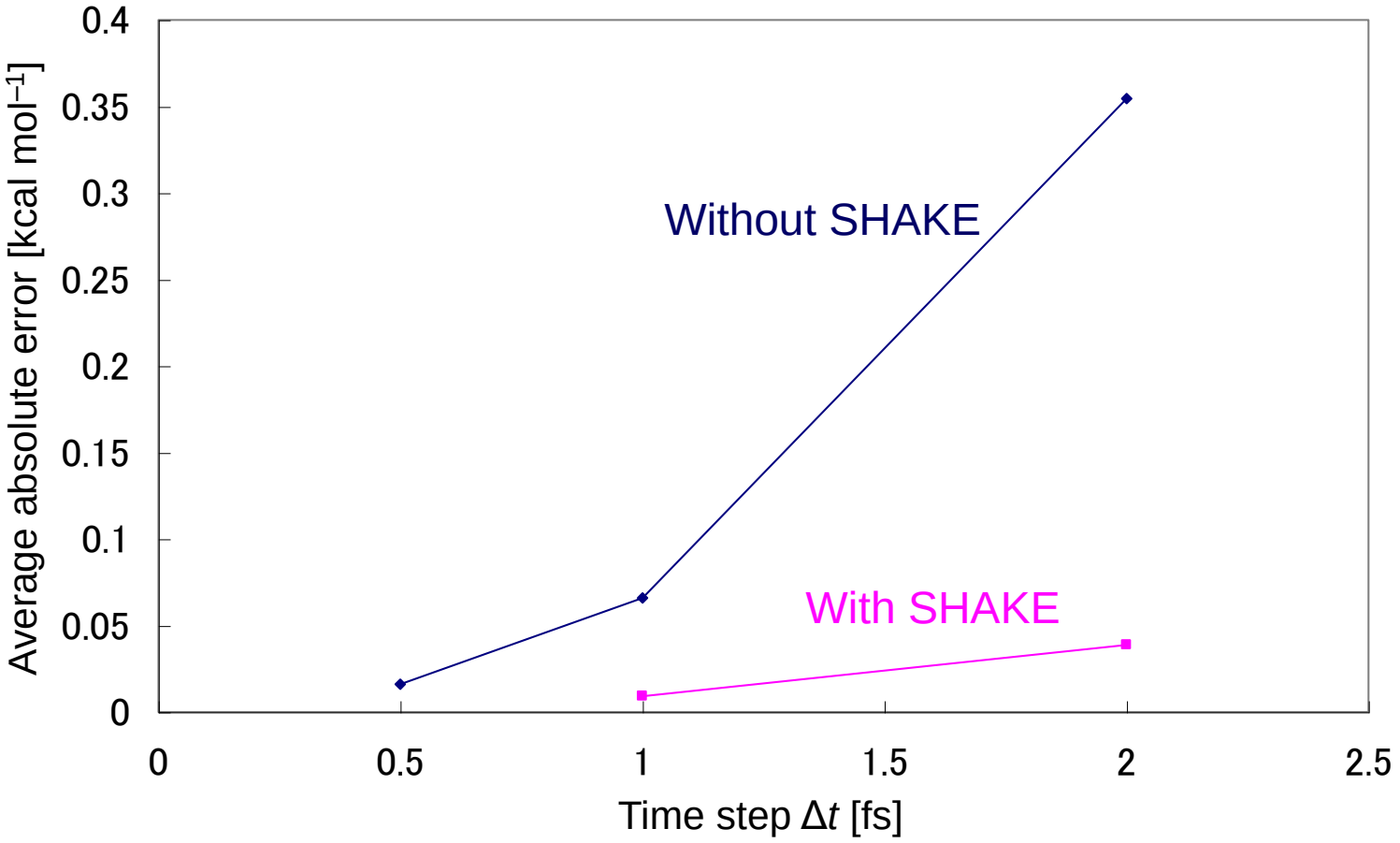
- In general, 1/10 – 1/20 of the cycle of the fastest motion is appropriate to the time step.
- X–H bond stretching motion is the fastest. →Cycle is about 10 fs→Δt = 0.5～1 fs
- X–X bond stretching motion is the second fastest.→Cycle is about 20 fs
- The SHAKE method fixes X–H bond lengths. →Longer time step (Δt = 2 fs) can be used.

An example (1)

Time evolutions of the error in the total energy during the constant-*NVE* MD simulations of methanol



An example (2)



The accuracy of the simulation with Δt = 2 fs and SHAKE is comparable to that of the simulation with Δt = 0.5 fs and without SHAKE.

Settings in NAMD (1)

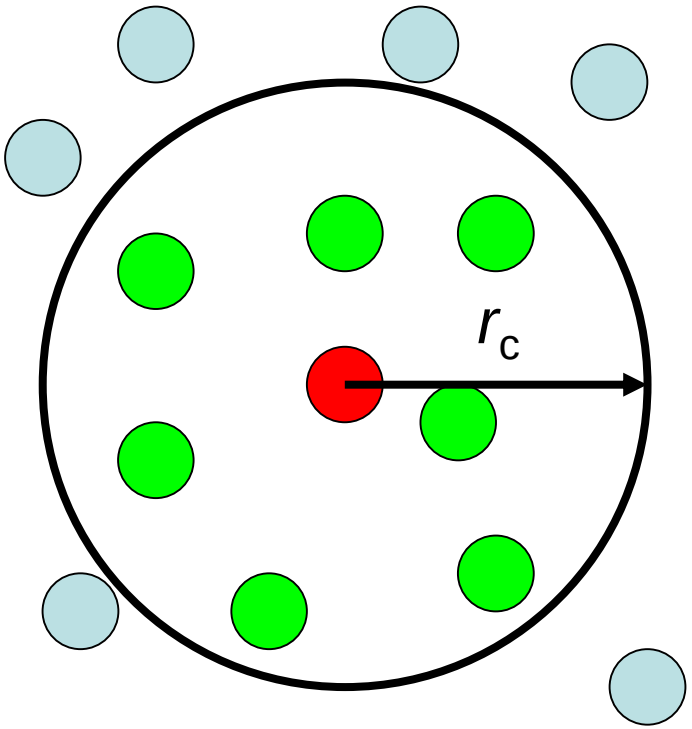
- SHAKE is enabled by the following settings.

```
rigidBonds          all
useGroupPressure    yes
```

Calculation of non-bonded interactions

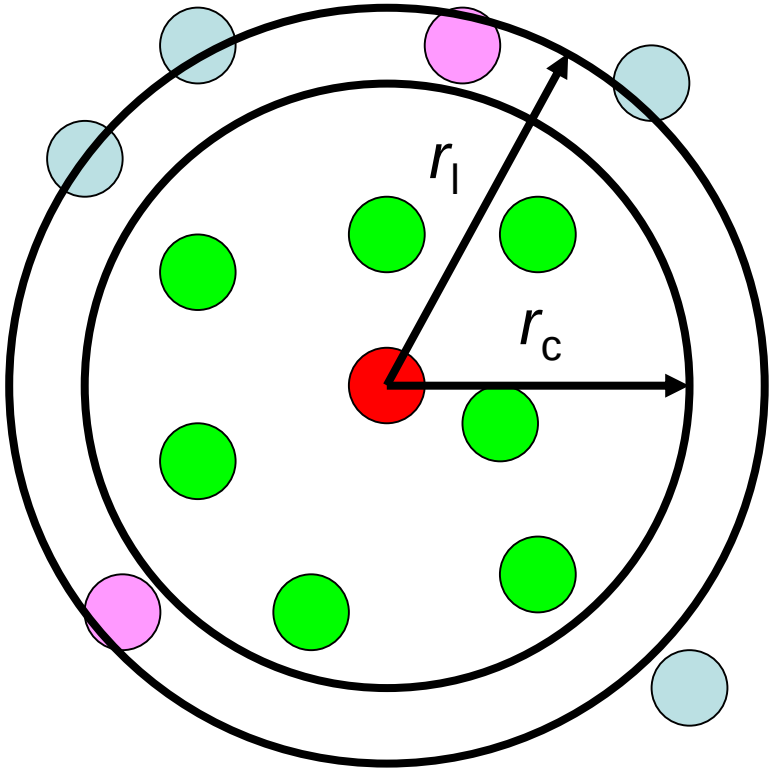
- Non-bonded interactions are calculated for pairs of atoms.
  $\rightarrow N(N-1)/2$ pairs in N-atom system
- Non-bonded interactions reduces as the distance between atoms increases. (Van der Waals attraction is proportional to r^{-6} and electrostatic interaction is to r^{-1}.)
- Interactions between distant atoms are negligible.→the cut-off method

The cut-off method



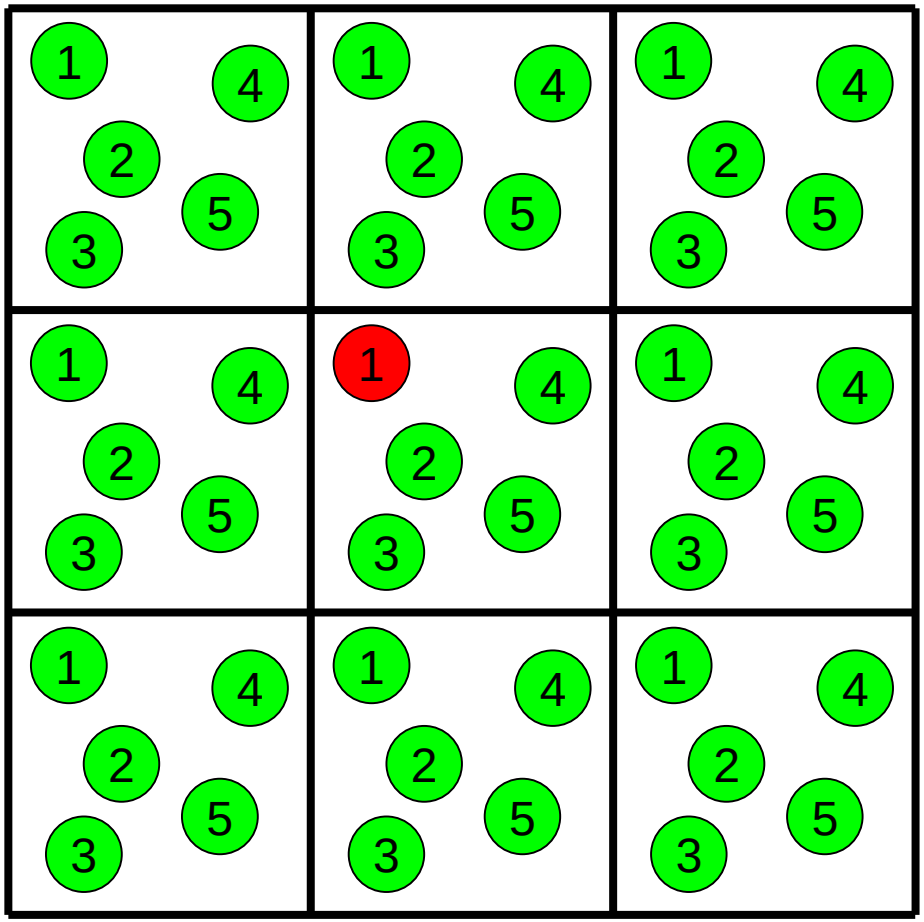
- Only the atoms within the radius of r_c from atom *i* interact with atom *i*.
- Let the average number of atoms within this sphere be *M*. Then, the number of pairs reduces to *NM* from $N(N-1)/2$.

Making pair list



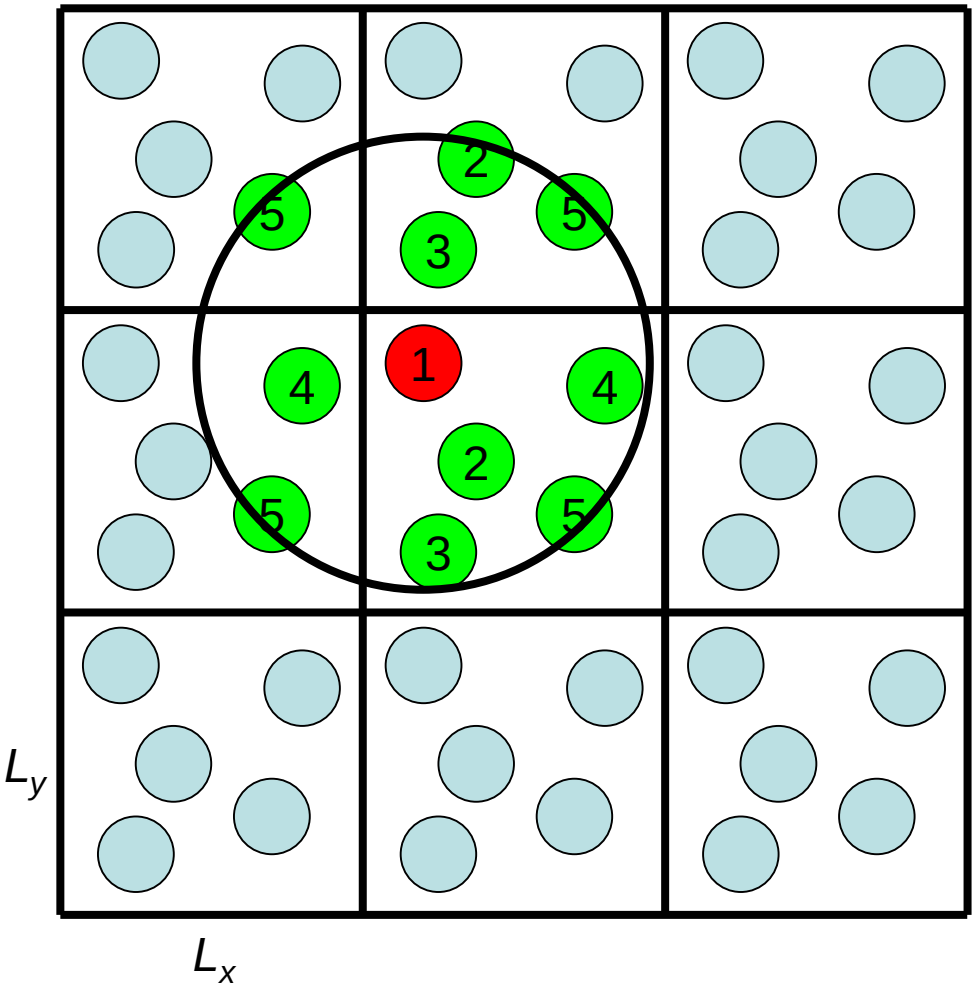
- It is necessary to list atoms within the cut-off radius r_c for each atom.
- Computational cost is proportional to $N(N-1)/2$.
- To reduce the cost, the list of atoms within the distance of r_l ($r_l > r_c$) for each atom is generated. The list is updated only when the displacement of an atom exceeds $r_l - r_c$.

Periodic boundary condition (1)



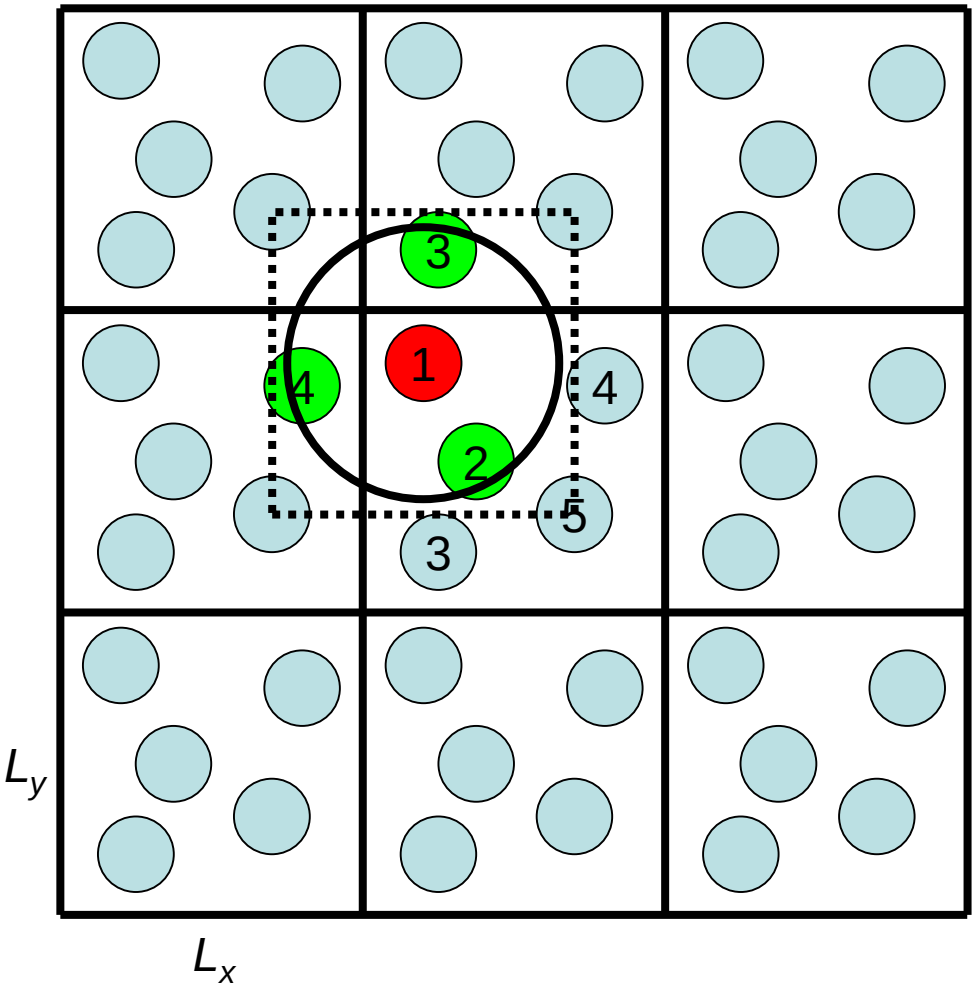
- Under the periodic boundary condition, it is impossible to compute all the interactions without any approximation.

Application of the cut-off method



- It is necessary to consider the atoms in the image cells neighboring to the central cell.
(In this case, $26N^2+N(N-1)/2$ pairs must be considered.)

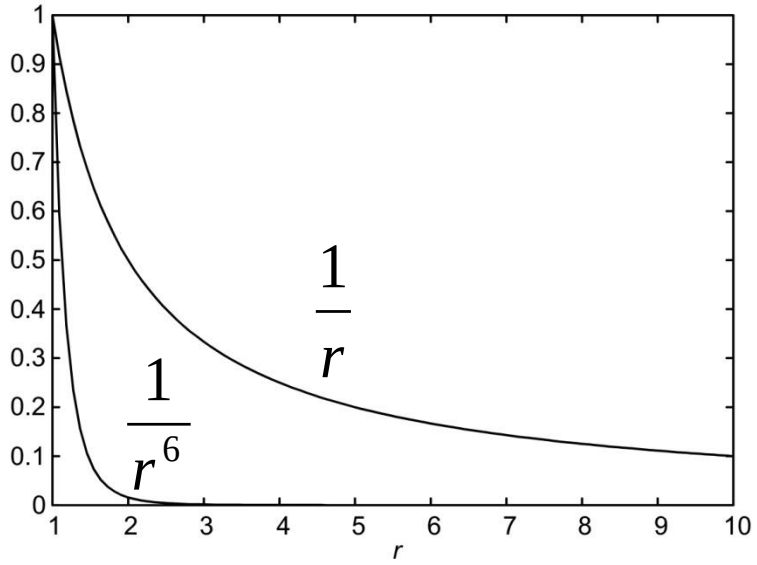
Minimum image convention

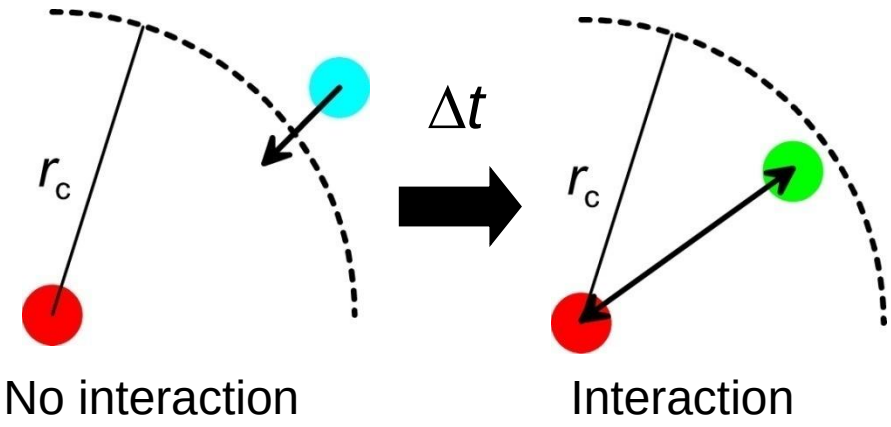


- When the cut-off radius is less than the half of the shortest dimension of the cell, the number of pairs that must be considered reduces to $N(N-1)/2$.
→minimum image convention

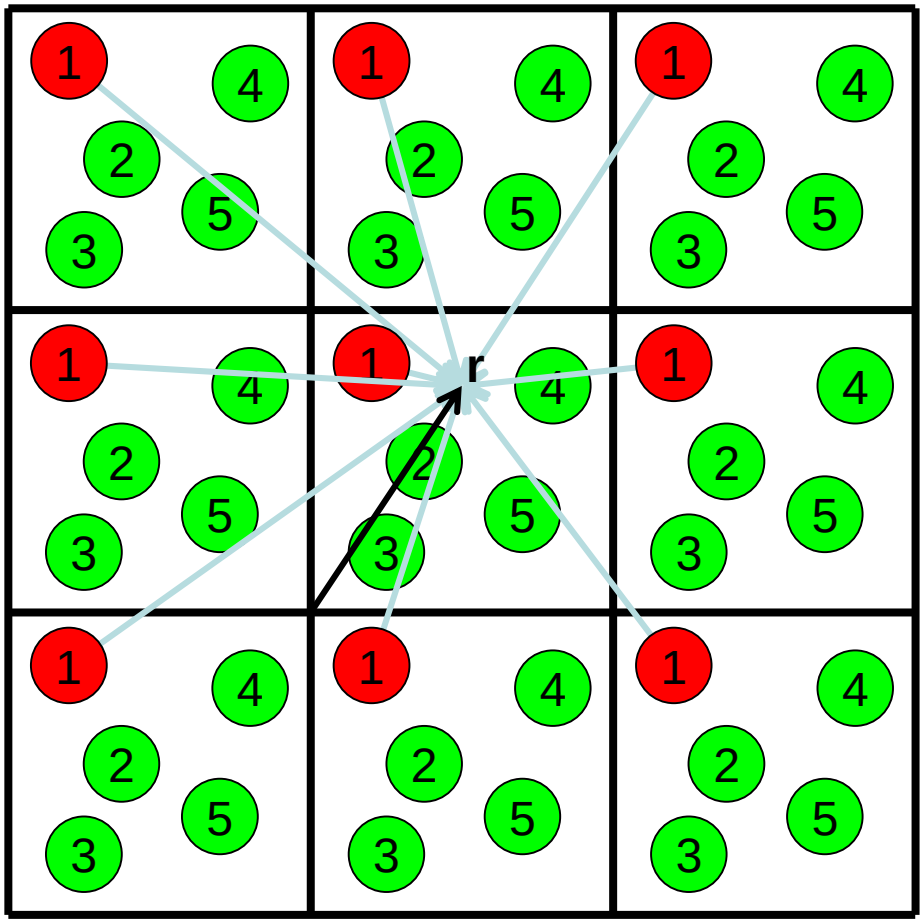
Problem of the cut-off method

- Van der Waals interaction is proportional to r^{-6}.
  →It can be accurately calculated with the cut-off method.
- Electrostatic interaction is proportional to r^{-1}.
  →The cut-off method causes large errors.
- Because the energy changes discontinuously when an atom goes in and out the cut-off sphere, the total energy is not conserved.





Calculation without cut-off (1)



Interactions are calculated not only between the atoms in the central cell, but also between the atoms of the central cell and those of the image cells.

Electrostatic potential at **r** is given by,

$$\varphi(\mathbf{r}) = \sum_{\mathbf{n}} \sum_{j} \frac{q_j}{\left|\mathbf{r} - \mathbf{r}_j + \mathbf{Ln}\right|}$$

Calculation without cut-off (2)

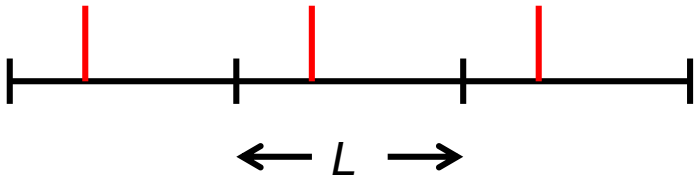
- The charge distribution is periodic.
  - The charge distribution can be expressed with a Fourier series.

$$\rho(x) = \sum_{n=-\infty}^{\infty} \tilde{\rho}_n \exp\left(2\pi i n \frac{x}{L}\right)$$

$$\tilde{\rho}_n = \frac{1}{L}\int_0^L \rho(x)\exp\left(-2\pi i n \frac{x}{L}\right)dx$$

$$\rho(x+L) = \sum_{n=-\infty}^{\infty} \tilde{\rho}_n \exp\left[2\pi i n \frac{(x+L)}{L}\right]$$

$$= \sum_{n=-\infty}^{\infty} \tilde{\rho}_n \exp\left(2\pi i n \frac{x}{L}\right) = \rho(x)$$



Calculation without cut-off (3)

- The 3D version:

$$\rho(\mathbf{r}) = \sum_{\mathbf{n}} \tilde{\rho}_{\mathbf{n}} \exp\left(2\pi i n_x \frac{x}{L_x}\right) \exp\left(2\pi i n_y \frac{y}{L_y}\right) \exp\left(2\pi i n_z \frac{z}{L_z}\right)$$

$$= \sum_{\mathbf{n}} \tilde{\rho}_{\mathbf{n}} \exp\left(2\pi i \mathbf{L}^{-1}\mathbf{n}\cdot\mathbf{r}\right), \quad \mathbf{L} = \begin{pmatrix} L_x & 0 & 0 \\ 0 & L_y & 0 \\ 0 & 0 & L_z \end{pmatrix}$$

$$\tilde{\rho}_{\mathbf{n}} = \frac{1}{|\mathbf{L}|} \int_{\text{cell}} \rho(\mathbf{r}) \exp\left(-2\pi i \mathbf{L}^{-1}\mathbf{n}\cdot\mathbf{r}\right) d\mathbf{r}, \quad |\mathbf{L}| = L_x L_y L_z$$

Calculation without cut-off (4)

- Solution of the Poisson equation

$$\nabla^2\varphi(\mathbf{r}) = -4\pi\rho(\mathbf{r})$$

$$\tilde{\rho}_{\mathbf{n}} = -\frac{1}{4\pi|\mathbf{L}|}\int_{\text{cell}} \nabla^2\varphi(\mathbf{r})\exp\left(-2\pi i\mathbf{L}^{-1}\mathbf{n}\cdot\mathbf{r}\right)d\mathbf{r}$$

$$= \frac{\pi\left|\mathbf{L}^{-1}\mathbf{n}\right|^2}{|\mathbf{L}|}\int_{\text{cell}} \varphi(\mathbf{r})\exp\left(-2\pi i\mathbf{L}^{-1}\mathbf{n}\cdot\mathbf{r}\right)d\mathbf{r}$$

$$= \pi\left|\mathbf{L}^{-1}\mathbf{n}\right|^2\tilde{\varphi}_{\mathbf{n}}$$

$$\varphi(\mathbf{r}) = \sum_{\mathbf{n}} \frac{\tilde{\rho}_{\mathbf{n}}}{\pi\left|\mathbf{L}^{-1}\mathbf{n}\right|^2}\exp\left(2\pi i\mathbf{L}^{-1}\mathbf{n}\cdot\mathbf{r}\right)$$

→Potential energy function

To void divergence, the net charge ($\tilde{\rho}_0$) must be zero.

The particle mesh Ewald method (1)

- n_x, n_y, and n_z run from minus infinity to infinity.
- By discretizing **r**, the ranges can be narrowed.

$$\rho_{\mathbf{k}} = \rho\left(\frac{k_x}{K_x}L_x, \frac{k_y}{K_y}L_y, \frac{k_z}{K_z}L_z\right)$$

$$= \sum_{n_x=0}^{K_x-1}\sum_{n_y=0}^{K_y-1}\sum_{n_z=0}^{K_z-1}\tilde{\rho}_{\mathbf{n}}\exp\left(2\pi i\mathbf{K}^{-1}\mathbf{n}\cdot\mathbf{k}\right)$$

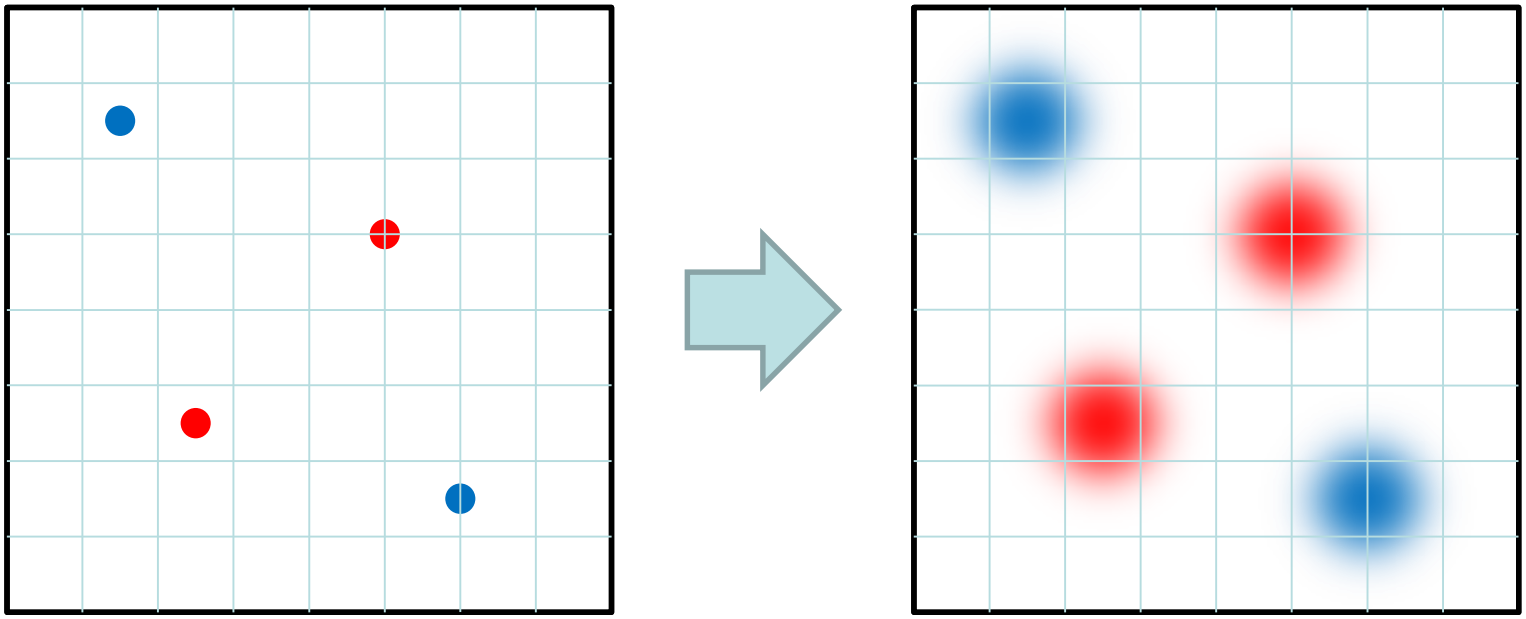
$$\tilde{\rho}_{\mathbf{n}} = \frac{1}{|\mathbf{K}|}\sum_{k=0}^{K-1}\rho_{\mathbf{k}}\exp\left(-2\pi i\mathbf{K}^{-1}\mathbf{n}\cdot\mathbf{k}\right)$$

$$\varphi_{\mathbf{k}} = \sum_{n_x=0}^{K_x-1}\sum_{n_y=0}^{K_y-1}\sum_{n_z=0}^{K_z-1}\frac{\tilde{\rho}_{\mathbf{n}}}{\pi\left|\mathbf{L}^{-1}\mathbf{n}\right|^2}\exp\left(2\pi i\mathbf{K}^{-1}\mathbf{n}\cdot\mathbf{k}\right)$$

Computed with the fast Fourier transform

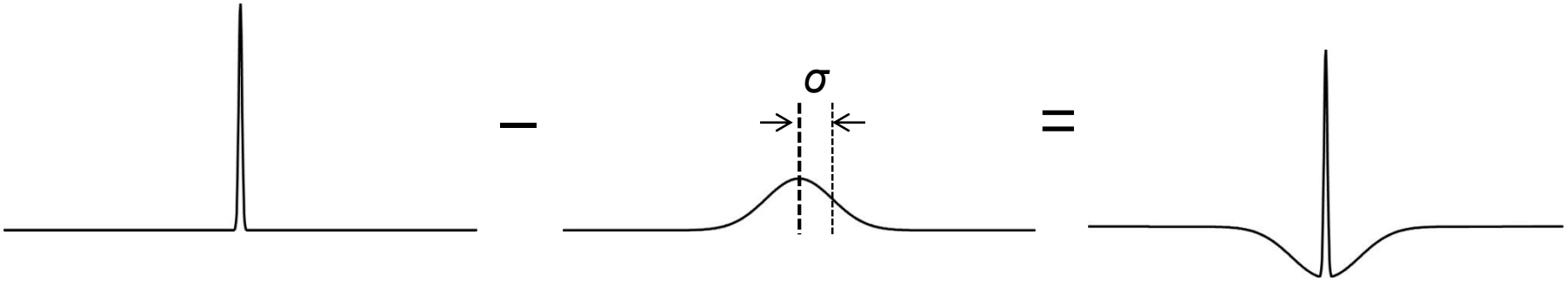
The particle mesh Ewald method (2)

- The charge distribution comprises point charges.
  → In general, the point charges are not on the lattice.
  →The distribution is convoluted with the Gaussian function.



The particle mesh Ewald method (3)

- Difference from the original charge distribution



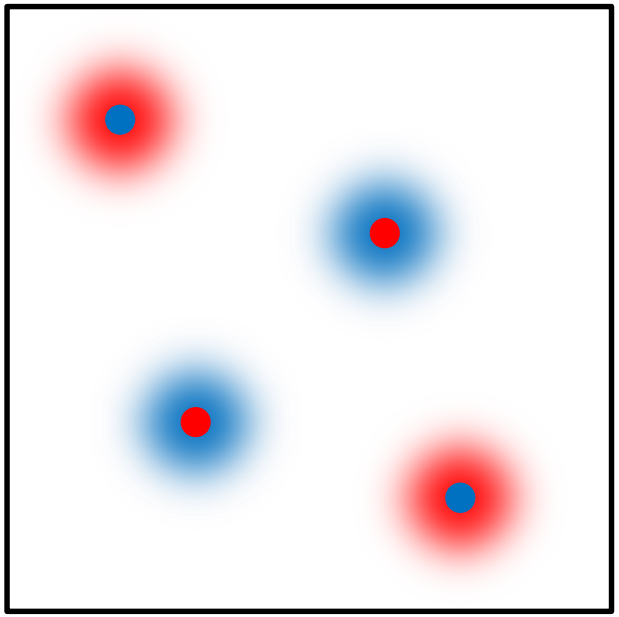
Point charge

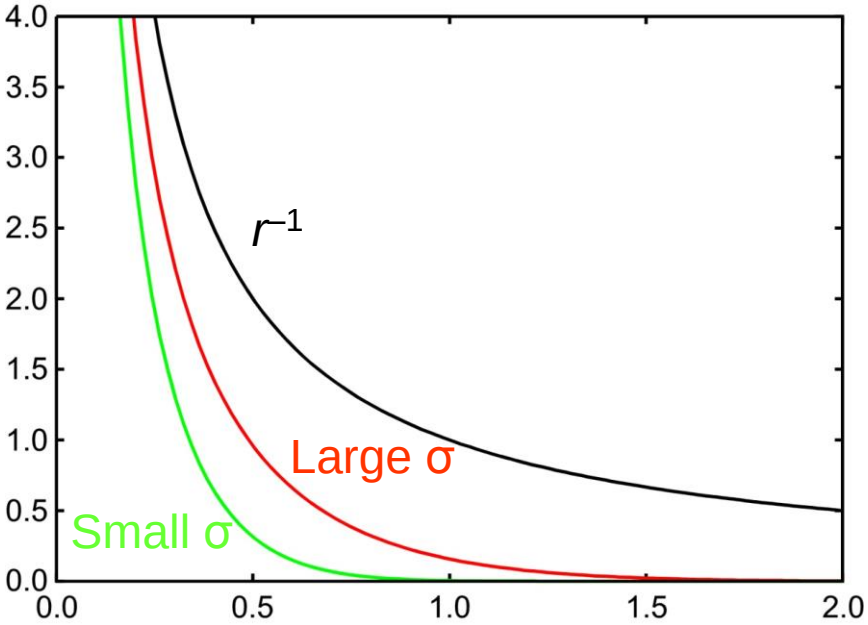
Convoluted distribution

Residual

- Electrostatic potential is expressed as the sum of the electrostatic potential produced by the convoluted charge distribution and that produced by the residual charge distribution.

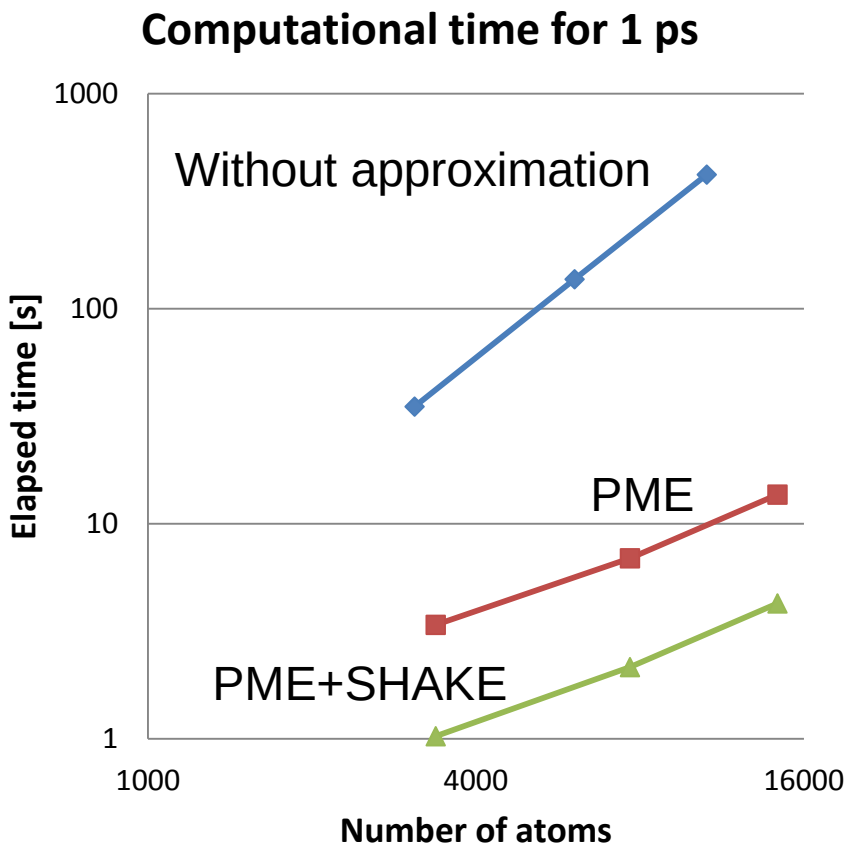
The particle mesh Ewald method (4)





- In the residual distributions, charges of opposite signs are distributed around the point charges.
  →Electrostatic potential more rapidly reduces than r^{-1}.
  →The interactions can be accurately calculated with the cut-off method.

Computational time (2)



- Simulations were performed on water boxes and spheres.
- Without approximation, computational time is proportional to N^2.
- With PME, it becomes proportional to $N\log N$.
- SHAKE allows the use of four-times longer time step (2 fs). The computational speed becomes 3.2-times faster.

Settings in NAMD (2)

- Use the following settings when using PME:

```
cutoff                  10.0
switching               off
cellBasisVector1 42.3810 0.0 0.0
cellBasisVector2 0.0 36.4706 0.0
cellBasisVector3 0.0 0.0 42.1148
PME                     yes
PMEGridSizeX            45
PMEGridSizeY            40
PMEGridSizeZ            45
extendedSystem          XSC_file_name
```

★ ☆

- Write either of ★ or ☆.

MD simulation of a protein (1)

1. Open a protein structure with PDB ID of 1CRN in Chimera.
2. Change the representation to sticks.
3. Add hydrogen atoms.
4. Solvate the protein in a rectangular box.
5. Add charges. Use “AMBER ff99SB” for the force field of standard residues.
6. Write parameters (prmtop) in files named “1CRN.” The force field is “AMBER ff99SB.”

MD simulation of a protein (2)

7. Save the coordinates in a PDB file named “1CRN.pdb.”
8. Download 1CRN.zip from the web page of this lecture, save it in the desktop, and unzip it.
9. Move 1CRN.prmtop, 1CRN.inpcrd, and 1CRN.pdb to the 1CRN folder.
10. Run restraint.pl to generate 1CRN_rest.pdb.
11. Modify the cell dimensions in min1.in.
12. Run run.bat to start the simulations. (They will take about 8 minutes in total.)

Assignment 2

- Plot the variation of the Cα RMSD.
  - Set Ignore hydrogens to “false.”
- Plot the temperature (TEMP) and potential (POTENTIAL) as a function of time during the equilibration runs (eq1 and eq2) and the production run (prod).
  - The time step Δt is 2 fs.
- What conclusion(s) can you draw from these plots?

Submitting assignments

- Put the results and discussion of the assignments in the slides of a PowerPoint file.
- Send the PowerPoint file as an attachment to email.
  - Do not send the Excel files. They are too large!
- Put “Molecular modeling assignments” in the Subject field of the email.
- Be sure to put your name and ID card number (if you are a student) in the body of the email.
- Send the email to Prof. Tohru Terada (tterada@iu.a.u-tokyo.ac.jp).