June 4, 2012
Molecular Modeling and Simulation

Applications of Molecular Dynamics

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

Contents

- Energy minimization of a peptide
- Modeling aqueous environment
- Molecular dynamics simulation of a peptide
  - Exercise 1
- Molecular dynamics simulation of a protein
  - Exercise 2
- Improvement of computational efficiency
- Preparations for simulation

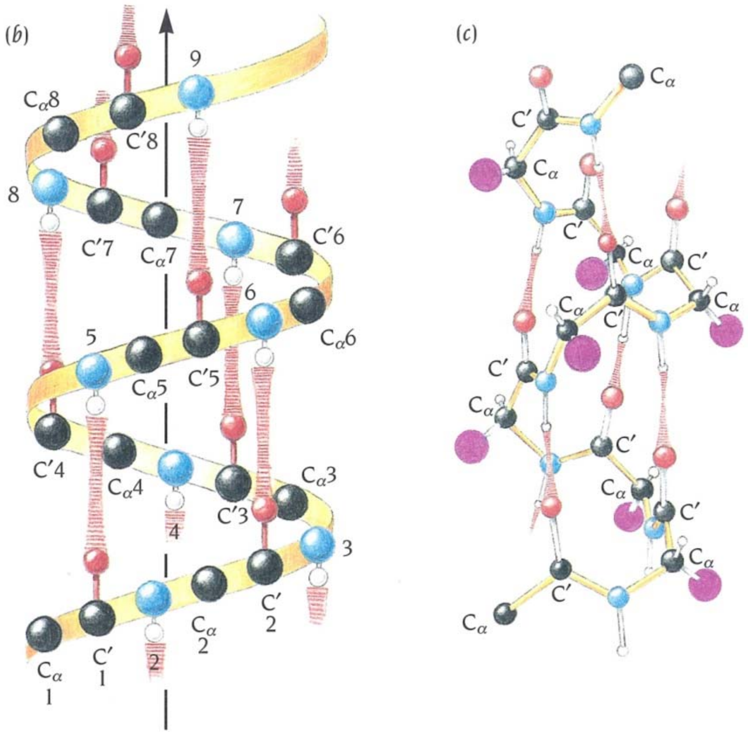
Building a peptide

1. Start Discovery Studio 3.0 Client and open a new Molecule Window
2. Click “Macromolecules” button
3. Expand “Build and Edit Protein”
4. Set Build Action to “Create/Grow Chain”
5. Set Conformation to “Right-hand Alpha Helix”
6. Click 9-times “Ala” in Choose Amino Acid to build an alanine 9-mer peptide.

Measuring interatomic distance

- Measure interatomic distances between backbone amide nitrogen and backbone carbonyl oxygen atoms that form hydrogen bonds as follows:

1. Select two atoms by clicking on them while holding the Shift Key.
2. Click “Measure” button. The atoms are connected by a green line and the distance is displayed.

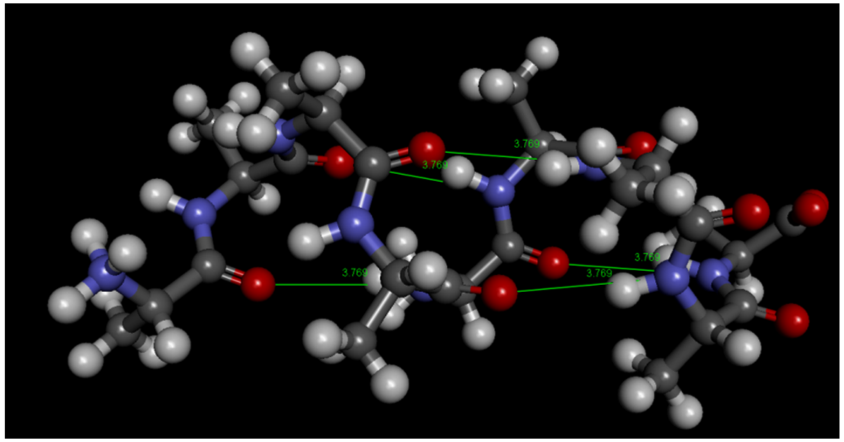


In an α helix, hydrogen bonds are formed between the carbonyl oxygen atom of the *i*-th residue and the amide nitrogen atom of the *i*+4-th residue.

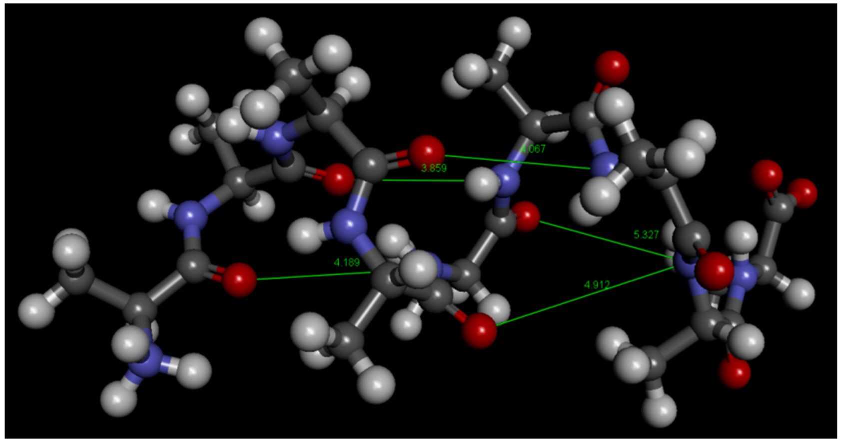
Energy minimization of a peptide

1. Save the peptide structure you have built as “ala9.dsv” on Desktop.
2. Click “Simulation” button.
3. Expand “Change Forcefield”, and set Forcefield to “charmm27.”
4. Click “Apply Forcefield.”
5. Expand “Run Simulations”, click “Minimization” in Tools section, and click “Run.”
6. After the job has been completed, check energies and hydrogen-bond distances in Data Table.

Result of energy minimization



Initial structure
268.307 kcal mol^{-1}



Energy-minimized structure
141.347 kcal mol^{-1}

Note that the hydrogen bonds are broken after the energy minimization.

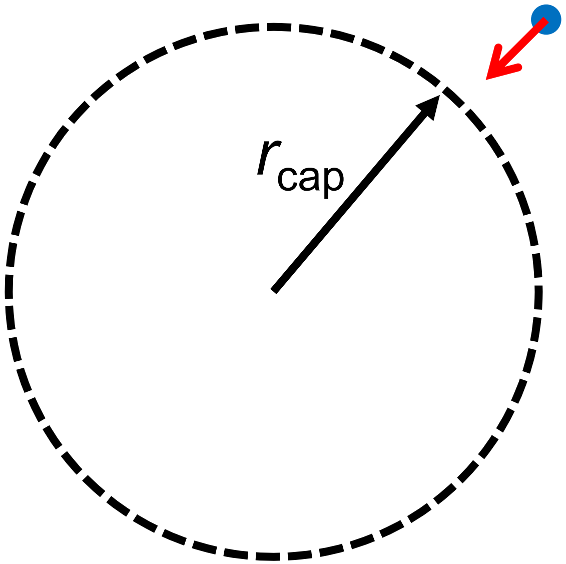
Modeling aqueous environment (1)

- This calculation is done in vacuum and the solvation effect of water is not considered.
- Simulations of biomacromolecules must be done in aqueous environment.

Modeling aqueous environment (2)

- Following models are usually used:
- Explicit water model
  - Arrange water molecules in a sphere shape
  - Arrange water molecules in a rectangular box →Periodic boundary condition
- Implicit solvent model
  - Non-polar effect→proportional to accessible surface area
  - Polar effect→Dielectric continuum model
    - Poisson-Boltzmann equation
    - Generalized Born model

Spherical arrangement



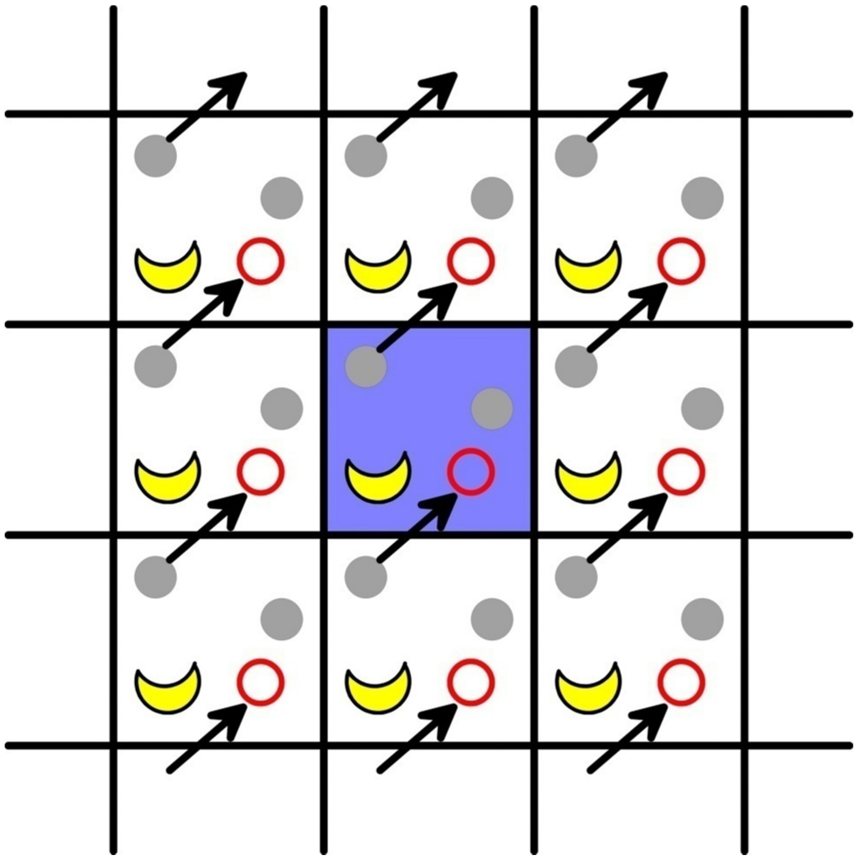
$$r = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}$$

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

- To avoid the water evaporation, restraining forces are imposed on the molecules that go outside the sphere.
- Water molecules near the surface are in an environment different from that of the water molecules near the center of the sphere.

Periodic 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 environment.
- The dimensions of the cell should be large enough to reduce the unnatural effect from neighboring cells.



Calculation of pressure

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

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

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

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

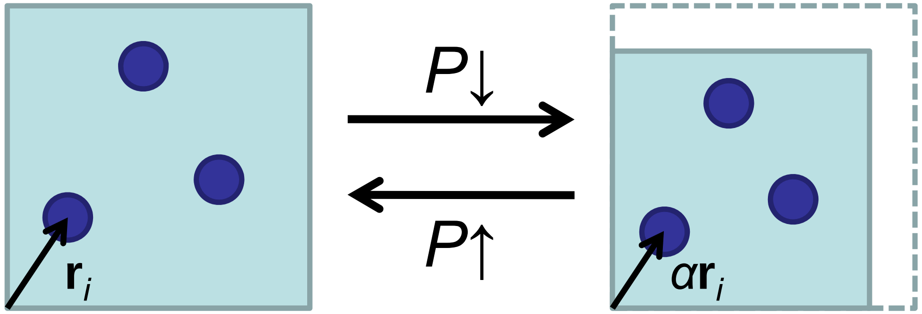
$$= \boxed{\frac{NkT}{V}} + \frac{1}{3V}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\mathbf{r}_{ij} \cdot \mathbf{f}_{ij}$$

Use this form under periodic boundary condition

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

Pressure regulation

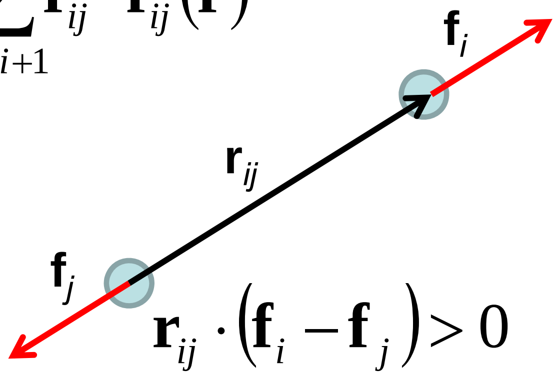
- Pressure is regulated by changing the dimensions of the cell.



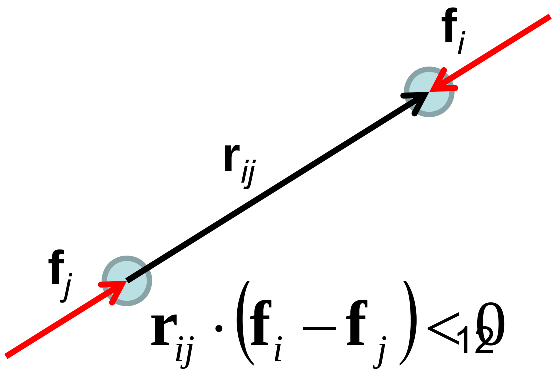
Center-of-mass positions of the molecules are scaled.

- Instantaneous pressure can assume negative values.

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



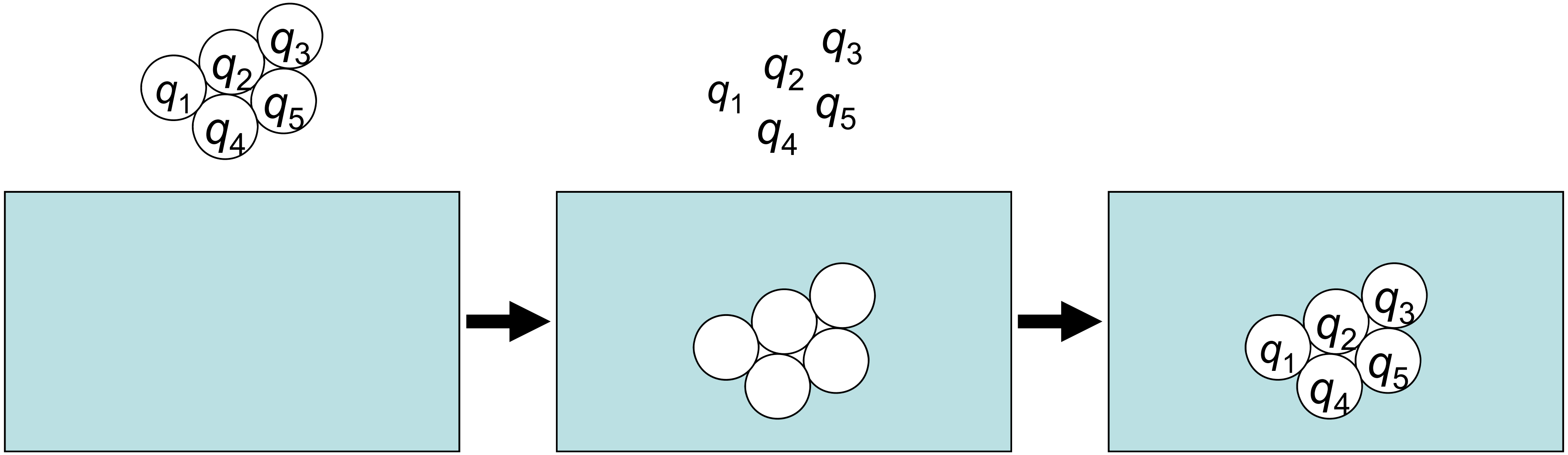
$$\mathbf{r}_{ij}\cdot(\mathbf{f}_i-\mathbf{f}_j)>0$$



$$\mathbf{r}_{ij}\cdot(\mathbf{f}_i-\mathbf{f}_j)<0$$

Implicit solvent model (1)

- Consider following thermodynamic process:



Solve uncharged solute: ΔG_{np}

Move the charges: ΔG_{pol}

- Free-energy change: $\Delta G_{\mathrm{solv}} = \Delta G_{\mathrm{np}} + \Delta G_{\mathrm{pol}}$

Implicit solvent model (2)

- Non-polar component (ΔG_{np}) can be estimated from an empirical relationship that ΔG_{np} is proportional to the solvent-accessible surface area (SASA).

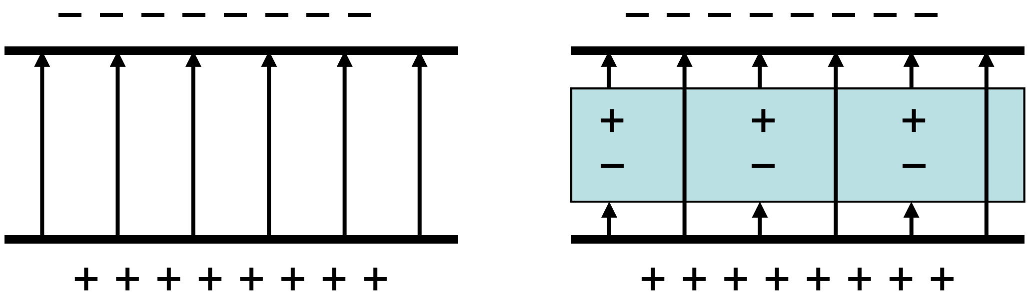
$$\Delta G_{\mathrm{np}} = \sigma A + b$$

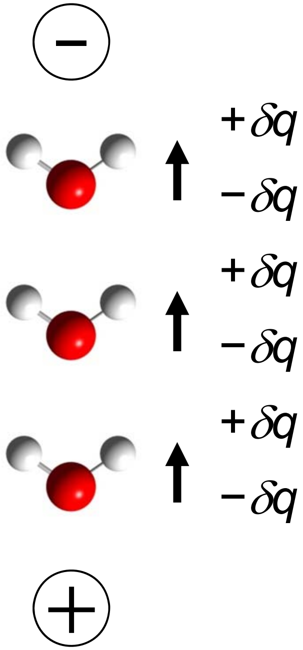
A: SASA、σ, b: empirical parameters

- Polar component (ΔG_{pol}) is calculated based on electromagnetism.

$$\Delta G_{\mathrm{pol}} = \frac{1}{2}\int \left[\underbrace{\phi(\mathbf{r})}_{\text{Electrostatic potential}} - \phi_{\mathrm{vac}}(\mathbf{r})\right]\underbrace{\rho(\mathbf{r})}_{\text{Charge distribution}}\, d\mathbf{r}$$

A dielectric



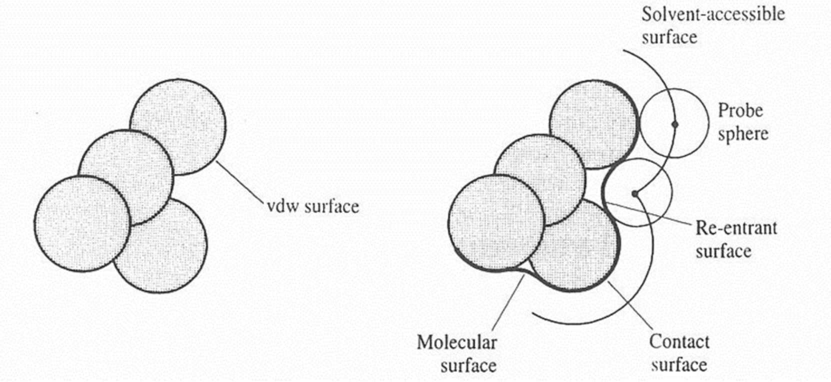


Charges induced on the surface of the dielectric partially cancel the electric field between the capacitor plates.
→Electrostatic potential reduces to 1/*ε.*

In an aqueous solution, water molecules work as a dielectric and weaken electrostatic interactions.

Dielectric continuum model

- Solvent accessible surface (SAS) is the trace of the center of the probe sphere (1.4 Å) rolling over the van der Waals surface of a molecule.
- Molecular surface (MS) is defined as a set of points that are in the distance of the probe radius from it.



- The inside of MS is filled with low (ε = 1～4) dielectric medium, whereas the outside is filled with high (ε = 80) dielectric medium.

Leach, A.R. “Molecular modeling – principles and applications” 2^{nd} Ed. Pearson, England

Poisson-Boltzmann (PB) equation

- PB equation gives electrostatic potential of a dielectric continuum model.
- Without salt→Poisson equation

$$\nabla \cdot \left[\varepsilon(\mathbf{r}) \nabla \underline{\phi(\mathbf{r})}\right] = -4\pi \underline{\rho(\mathbf{r})}$$

Electrostatic potential   Charge distribution of solute

- With salt→Distribution of the ions of the salt follows the Boltzmann distribution.

$$\nabla \cdot \left[\varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r})\right] = -4\pi \left[\rho(\mathbf{r}) + \underline{\rho_{\mathrm{ion}}(\mathbf{r})}\right]$$

Charge distribution of ions

Generalized Born (GB) model

- Problem of Poisson-Boltzmann equation
  - Force cannot be calculated analytically.
  - High computational cost
- Generalized Born model
  - Extends the Born equation that calculate solvation free-energy of an ion.

$$\Delta G_{\mathrm{pol}} = -\frac{q^2}{2a}\left(1 - \frac{1}{\varepsilon}\right) \Rightarrow \Delta G_{\mathrm{pol}} = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right)\sum_{i,j=1}^{N}\frac{q_i q_j}{f_{\mathrm{GB}}}$$

$$f_{\mathrm{GB}} = \sqrt{r_{ij}^2 + a_i a_j \exp\left(-r_{ij}^2 / 4a_i a_j\right)}$$

  - Low computational cost
  - Force can be calculated analytically.

Non-polar component

- Logarithm of molar solubility (*i.e.* ΔG_{np}) of hydrocarbon is proportional to its solvent-accessible surface area (SASA).*
- Standard value for parameter σ is 5 cal mol^{-1} Å^{-2}.**
- When GB or PB model is used to calculate the polar component, the combination with this model is called GB/SA or PB/SA model.

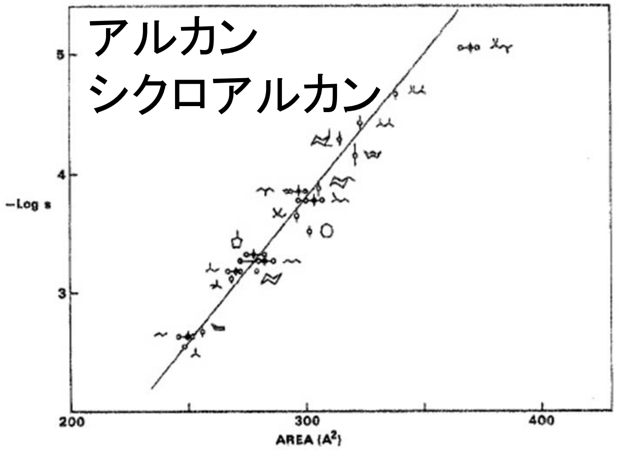


Figure 2. Relationship of cavity surface area to solubility for alkanes and cycloalkanes.

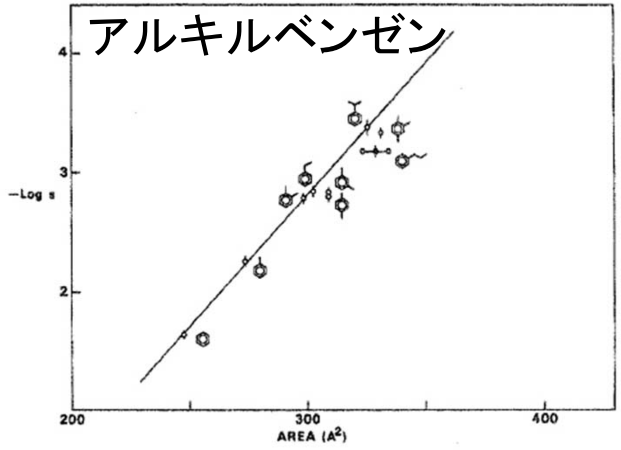


Figure 3. Relationship of cavity surface area to solubility for alkylbenzenes.

*Hermann *J. Phys. Chem.* **76**, 2754 (1972).

Sitkoff *et al.* *J. Phys. Chem.* **98, 1978 (1994).

Settings in Discovery Studio (1)

- Explicit solvent models
  1. Click “Simulation” button, expand “Run Simulations,” and click “Solvation” in Advanced section.
  2. Set Solvation Model to “Explicit Periodic Boundary” or “Explicit Spherical Boundary with Harmonic Restraint” and click “Run.”
- Generalized Born models
  1. Set Implicit Solvent Model to “Generalized Born with a simple SWitching (GBSW).”

MD simulation of a peptide (1)

1. Open “ala9.dsv” that you saved previously.
2. Solvate the peptide in a water sphere.
3. Solvated peptide is displayed in a new Molecule Window. Apply charmm27 force field to it.
4. Expand “Run Simulations” and click “Standard Dynamics Cascade” in Dynamics section.
5. Set Electrostatics to “Spherical Cutoff” and click “Run.”

- Two energy minimizations and heating, equilibration, and production MD runs are executed sequentially.
- When an error occurs, clear force field, apply charmm22, clear force field again, and apply charmm27 again.

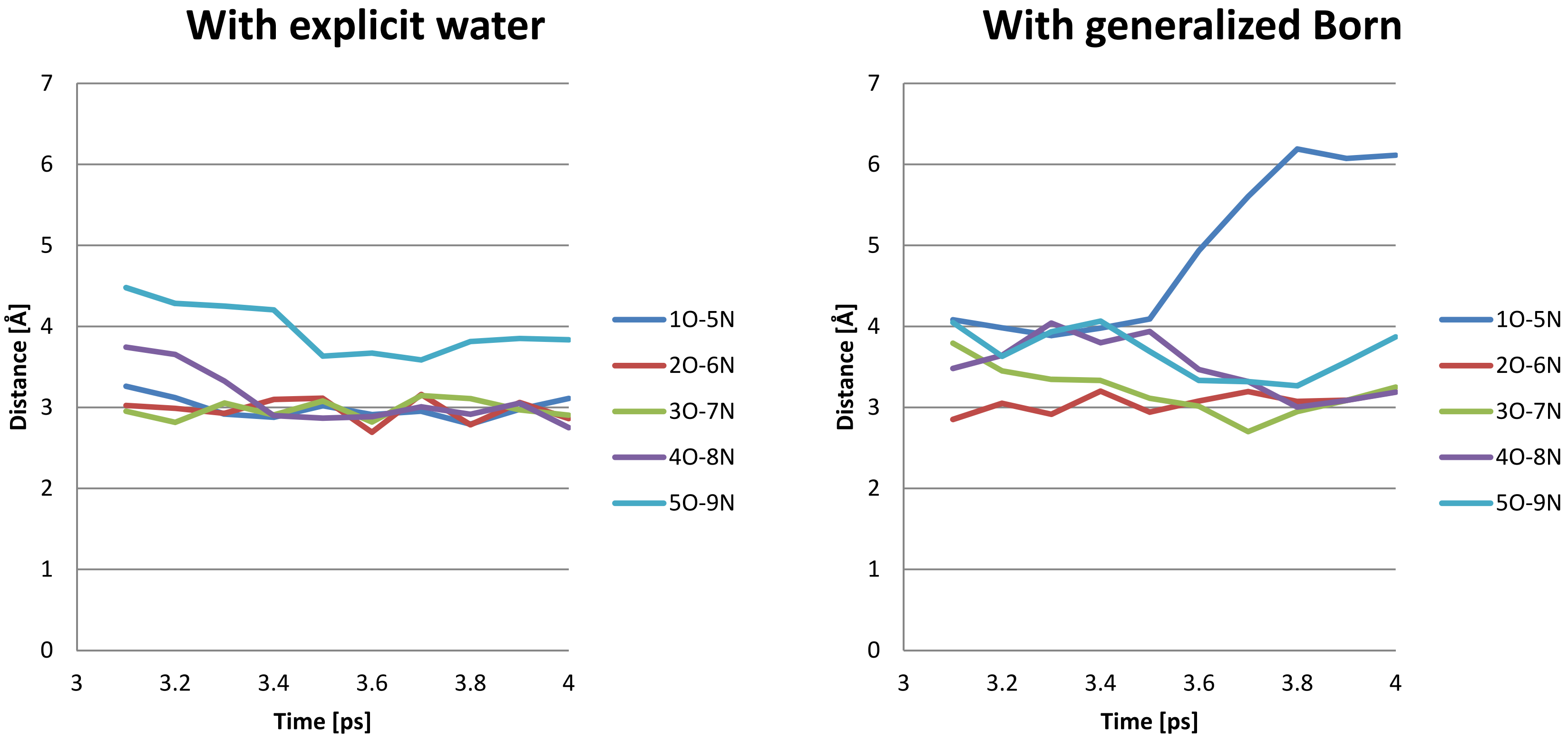
MD simulation of a peptide (2)

6. When the calculation has finished, a new Molecule Window appears and the result is shown in it.
7. Choose “View”→“Toolbars”→“Animation” from the menu.→Animation Toolbar is displayed.
8. Click “Start Animation” button ▶ to playback Animation.
9. Expand “Analyze Trajectory”, click “Analyze Trajectory” in Analyze section, and click “Run.”
10. Hydrogen bond distances are shown in the Conformation tab of Data Table.

Exercise 1

- Plot time evolution of the hydrogen bond distances of alanine 9-mer during the MD simulation in explicit water.
  - Select Time and Distance columns by clicking the header cells while holding the Ctrl key. Copy them and paste them into Excel.
  - Clarify between which atoms each distance is measured.
- Perform an MD simulation using generalized Born model (GBSW) and plot time variations of the hydrogen bond distances during the simulation.
- Briefly discuss the results.

Hydrogen bond distances



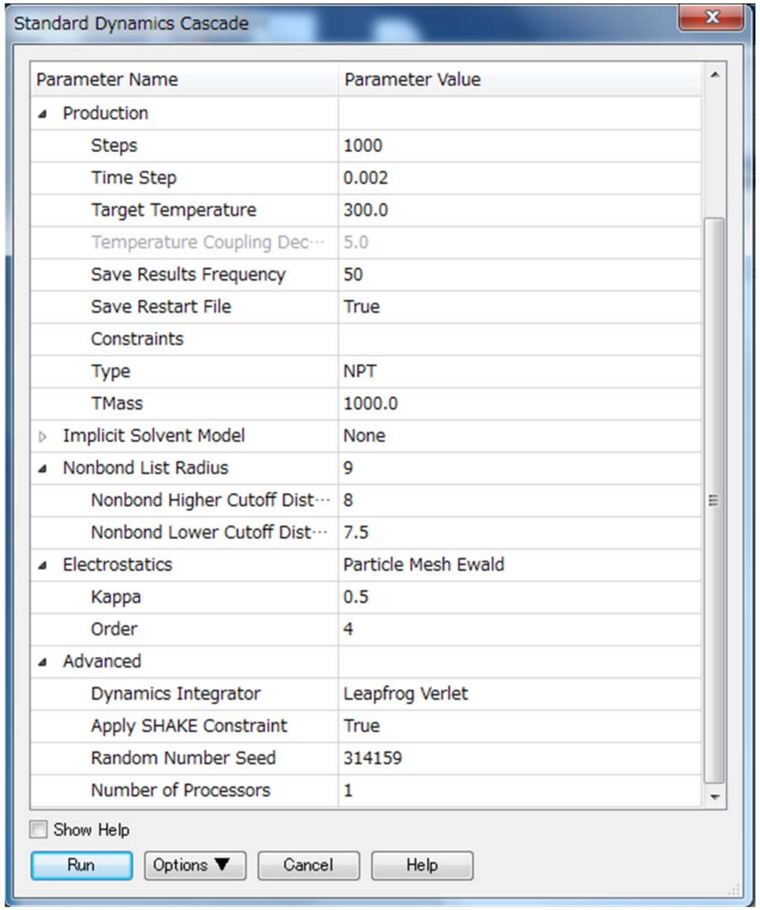
- Hydrogen bonds are almost maintained.
- In the MD with generalized Born model, the variations of the distances are larger.

MD simulation of a protein (1)

1. Open the structure of crambin (PDB ID: 1CRN).
2. Change Display Style to Line
3. Apply charmm27 force field. (Hydrogen atoms are added.)
4. Perform energy minimization with implicit solvent model, GBSW.
5. Solvate the protein in a rectangular water box.
6. Apply charmm27 force field again.

MD simulation of a protein (2)

7. Set the parameters of “Standard Dynamics Cascade” as shown here and click “Run.”
   - Steps: 1000
   - Time step: 2 fs
   - Constant-*NPT*
   - Particle Mesh Ewald
   - SHAKE



Parameter Name	Parameter Value
◢ Production	
Steps	1000
Time Step	0.002
Target Temperature	300.0
Temperature Coupling Dec···	5.0
Save Results Frequency	50
Save Restart File	True
Constraints	
Type	NPT
TMass	1000.0
▷ Implicit Solvent Model	None
◢ Nonbond List Radius	9
Nonbond Higher Cutoff Dist···	8
Nonbond Lower Cutoff Dist···	7.5
◢ Electrostatics	Particle Mesh Ewald
Kappa	0.5
Order	4
◢ Advanced	
Dynamics Integrator	Leapfrog Verlet
Apply SHAKE Constraint	True
Random Number Seed	314159
Number of Processors	1

Exercise 2

- Download “1CRN_10ps.dsv” from this lecture’s web page and open it. This file contains the trajectory of 10-ps simulation.
- Find more than two hydrogen bonds in the crystal structure of 1CRN.
- Plot the time evolution of the hydrogen bond distances during the MD simulation.
- Briefly discuss the results.

Computational time (1)

- Simulations were performed on water spheres.
- Sander module of Amber 11 was used.
- 8 cores of an Intel Xeon Processor were used.
- Time step (Δt) was 0.5 fs.
- Elapsed time for 1-ps simulation was measured.

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

*4.9 days / ns

Improving the efficiency

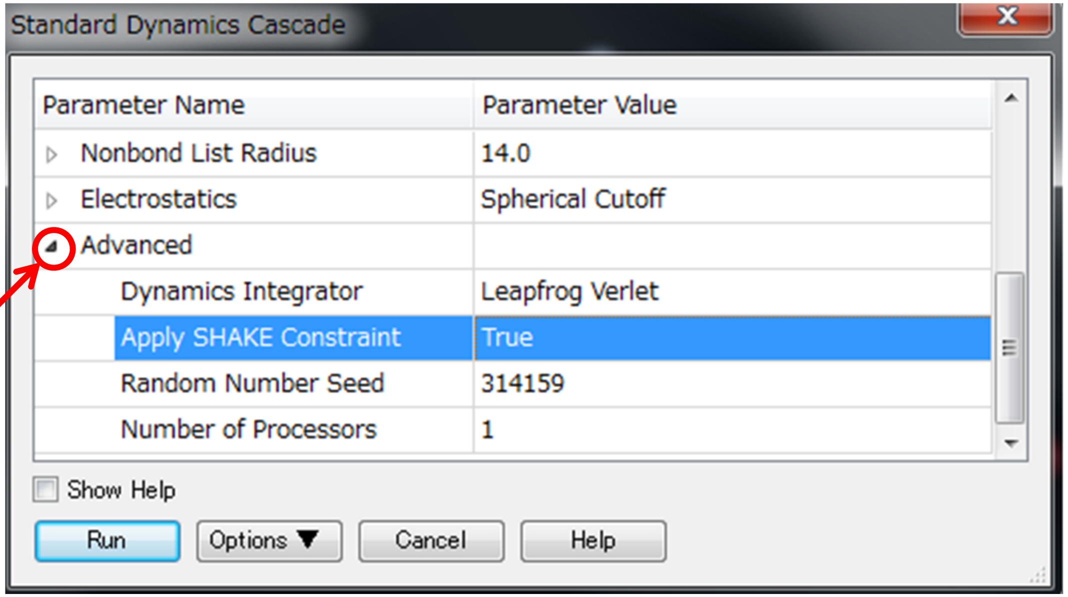
- Extend the time step
  - SHAKE method
  - Multiple time-step method
- Approximate the non-bonded interactions
  - Cut-off method
  - Multipole expansion
  - Particle mesh Ewald method

The SHAKE method

- In general, 1/10 – 1/20 of the cycle of the fastest motion is appropriate for 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.

Settings in Discovery Studio (2)

- SHAKE is not applied by default.
- To apply the SHAKE method, set the parameters of Standard Dynamics Cascade as shown here.



Standard Dynamics Cascade

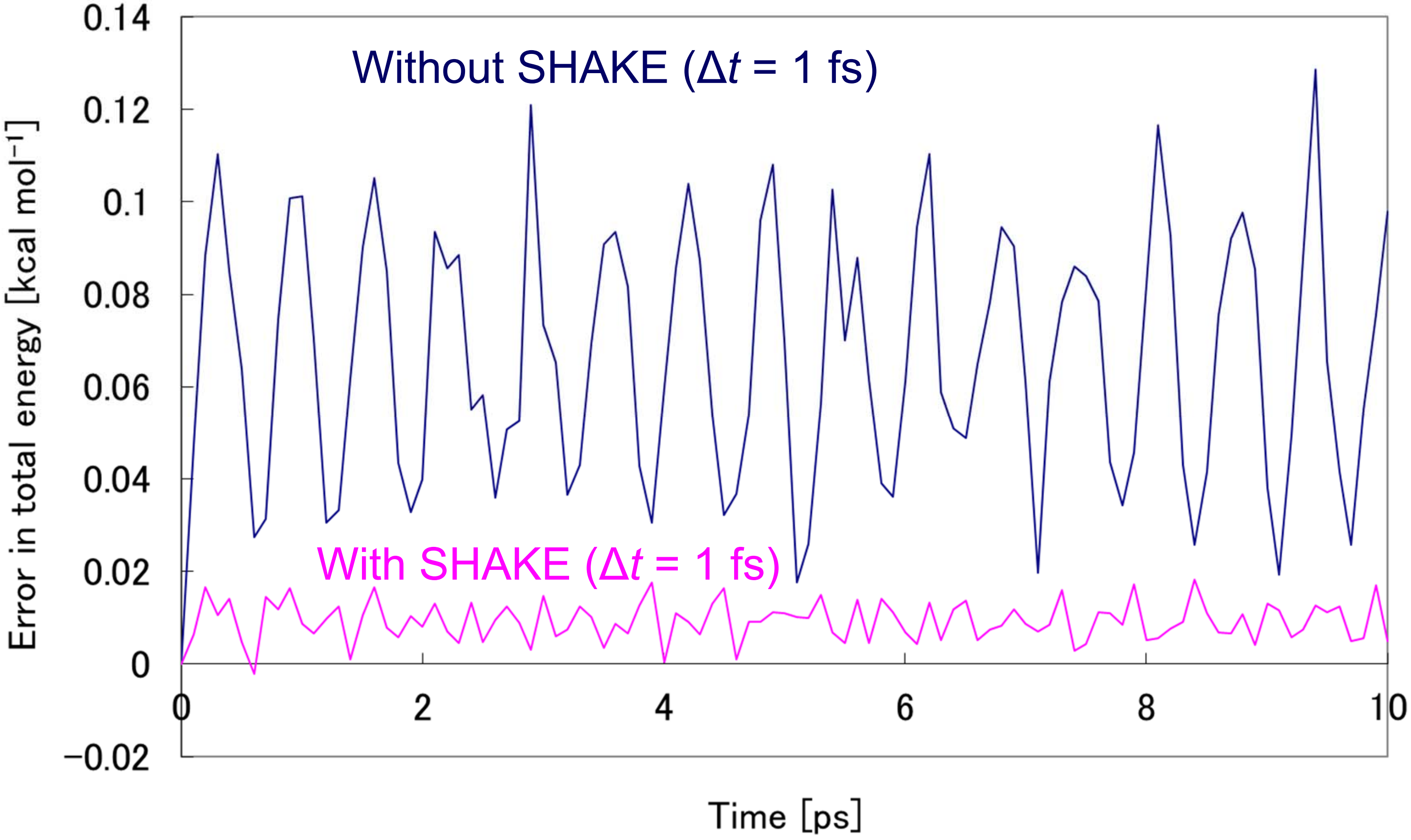
Parameter Name	Parameter Value
▷ Nonbond List Radius	14.0
▷ Electrostatics	Spherical Cutoff
◢ Advanced	
Dynamics Integrator	Leapfrog Verlet
Apply SHAKE Constraint	True
Random Number Seed	314159
Number of Processors	1

☐ Show Help

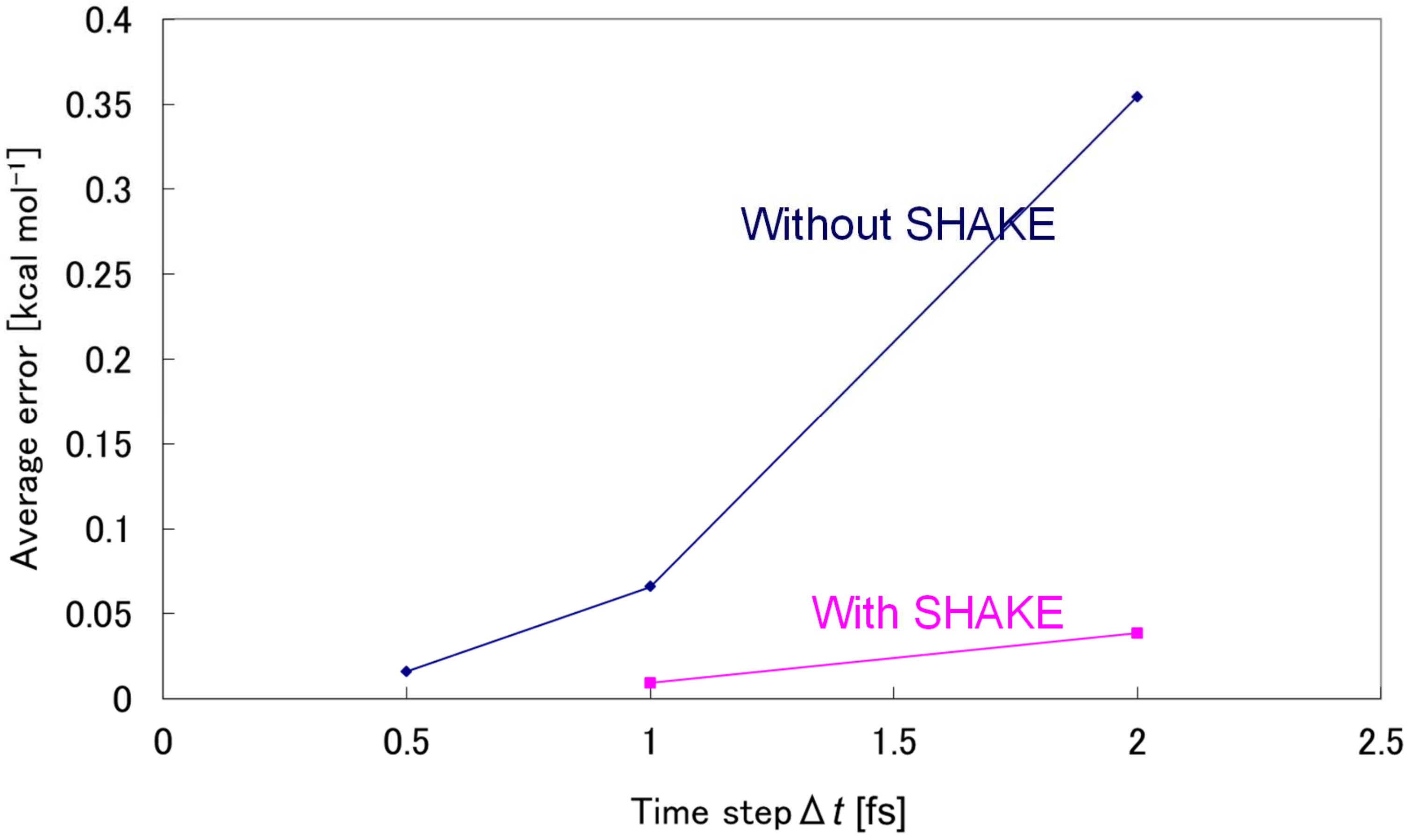
Click to expand

Effect of SHAKE (1)

Time evolution of the errors in the total energies during the constant-energy MD simulations of methanol



Effect of SHAKE (2)

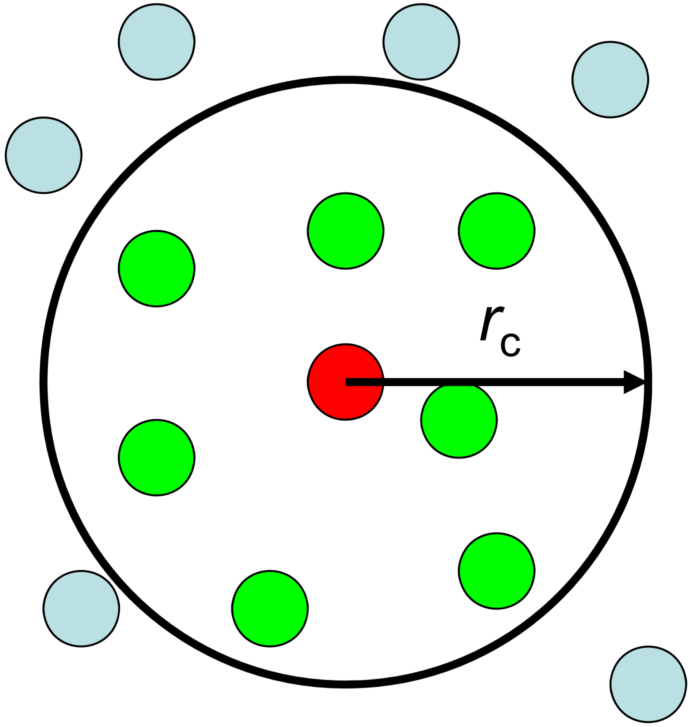


The accuracy of the simulation with Δ*t* = 2 fs and SHAKE is comparable to that of the simulation with Δ*t* = 0.5 fs and without SHAKE.

Calculation of non-bonded interactions

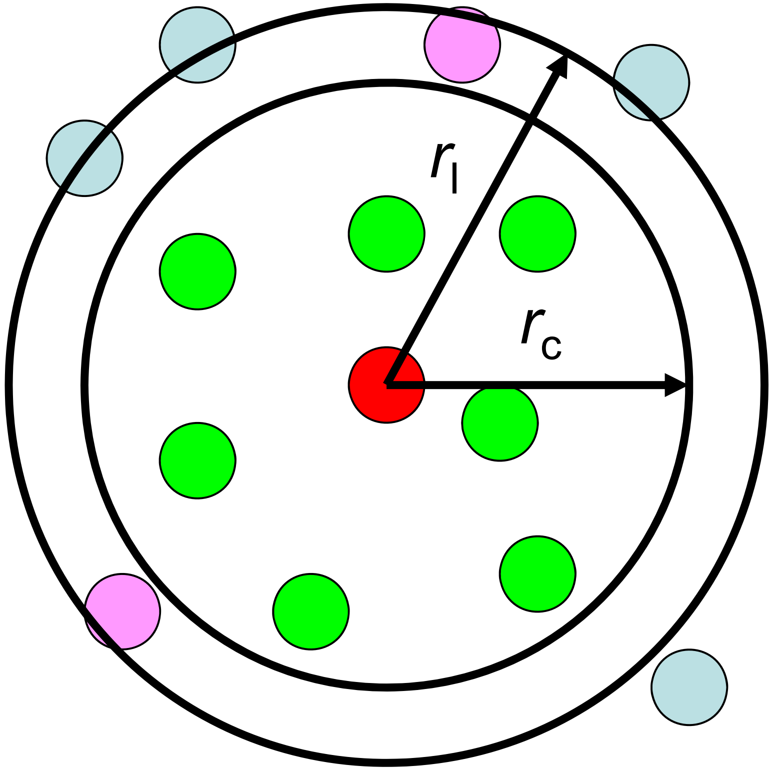
- Non-bonded interactions are calculated for pairs of atoms.
  →$N(N-1)/2$ pairs in N-atom system
- Non-bonded interactions reduces as the distance between atoms increases. (Van der Waals attraction is proportional to r^{-6} and electrostatic interaction is to r^{-1}.)
- Interactions between distant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

Cut-off method



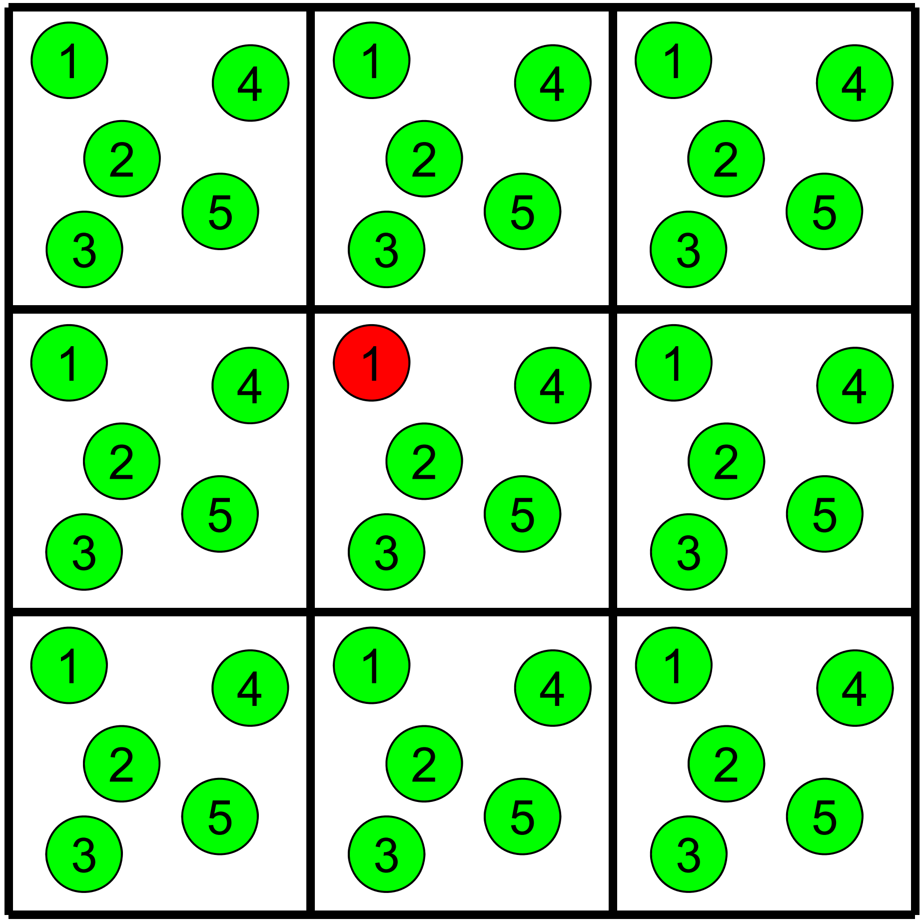
- Only the atoms within the radius of r_c from atom i interact with atom i.
- Letting the average number of atoms within this sphere be M, the number of pairs reduces to NM from $N(N-1)/2$.

Pair list



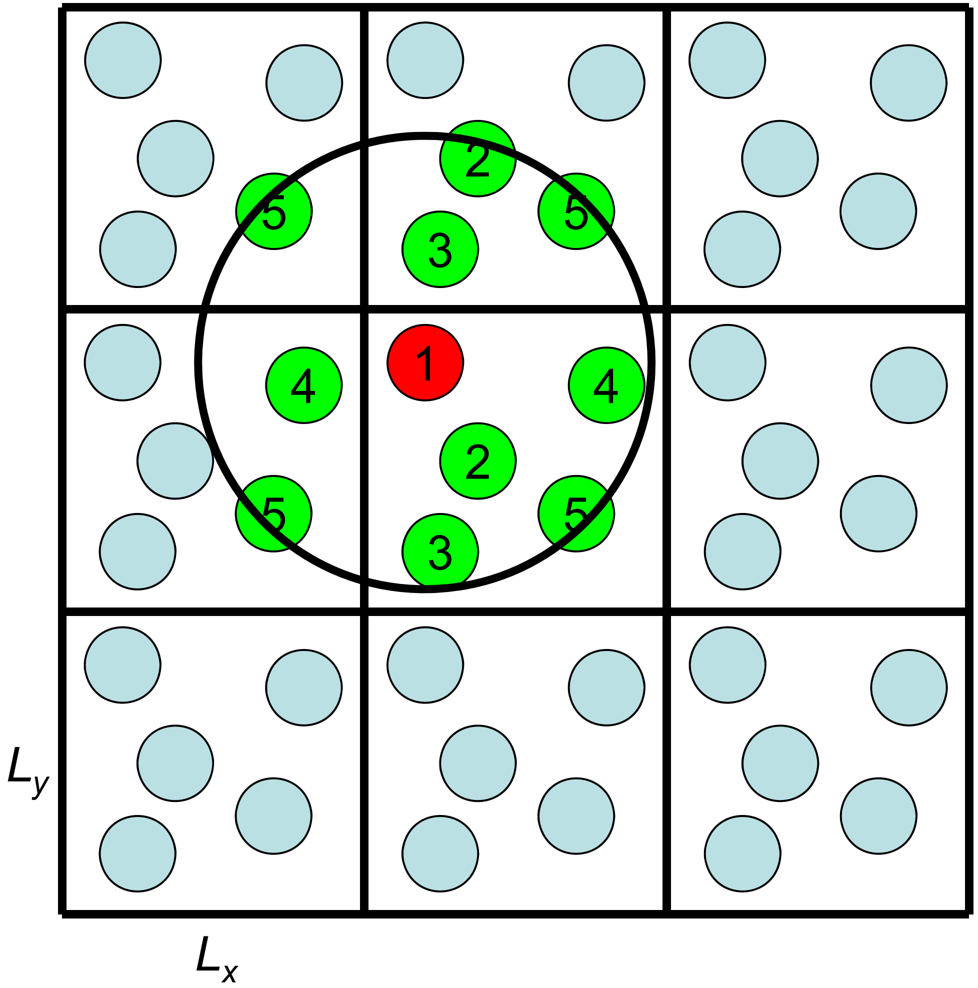
- It is necessary to list atoms within the cut-off radius r_c for each atom.
- Computational cost is proportional to $N(N-1)/2$.
- The list of atoms within the distance of r_l ($r_l > r_c$) for each atom is generated. The list is updated only when the displacement of an atom exceeds $r_l - r_c$.

Periodic boundary condition (1)



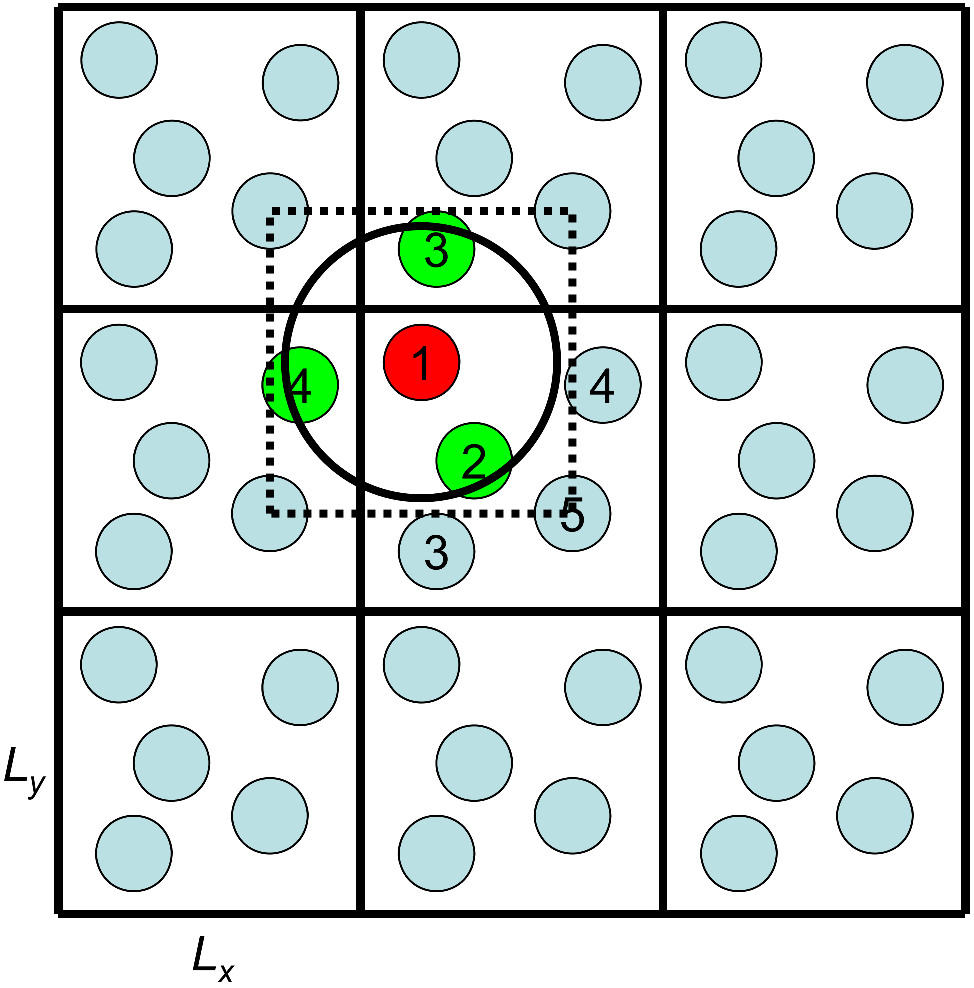
- Under the periodic boundary condition, it is impossible to compute all the interactions without any approximation.

Application of cut-off method



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

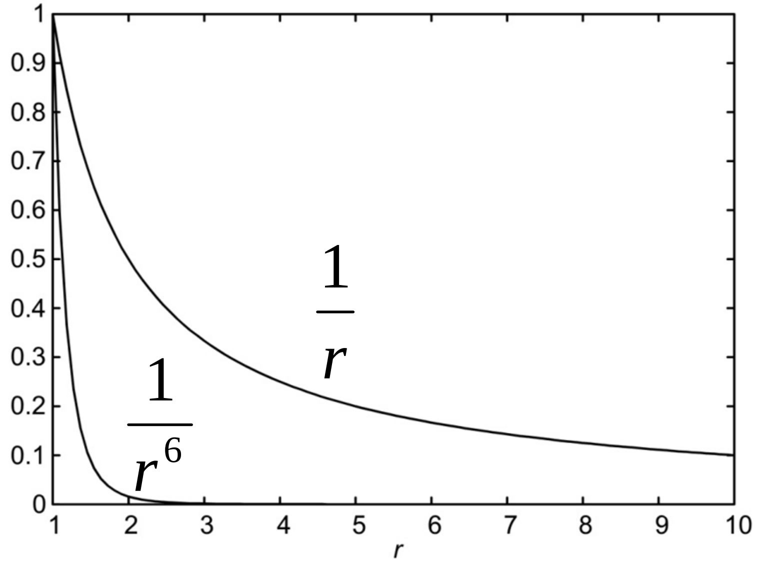
Minimum image convention

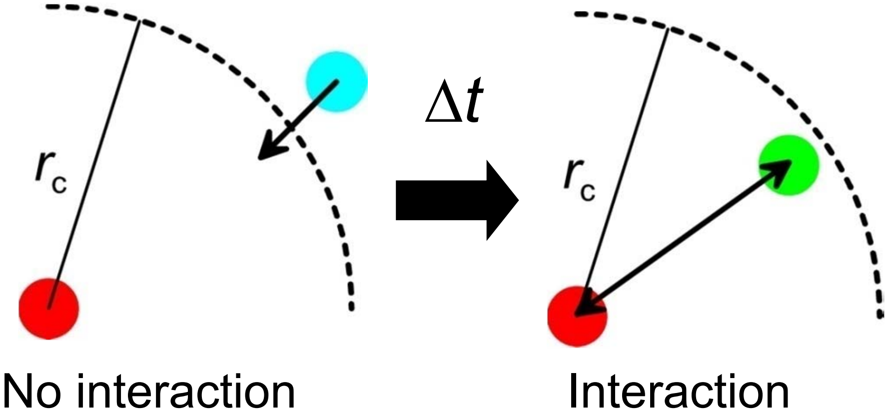


- When the cut-off radius is less than the half of the shortest dimension of the cell, the number of pairs that must be considered reduces to $N(N-1)/2$.
  →minimum image convention

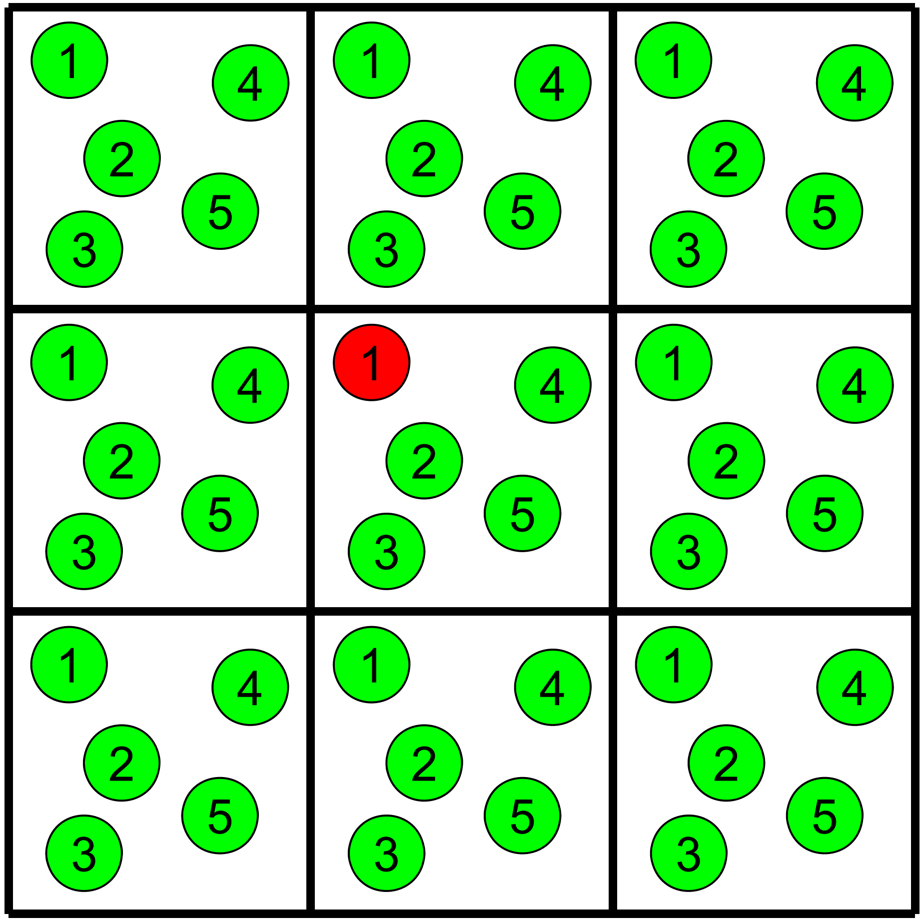
Problem of cut-off method

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





No cut-off method



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.

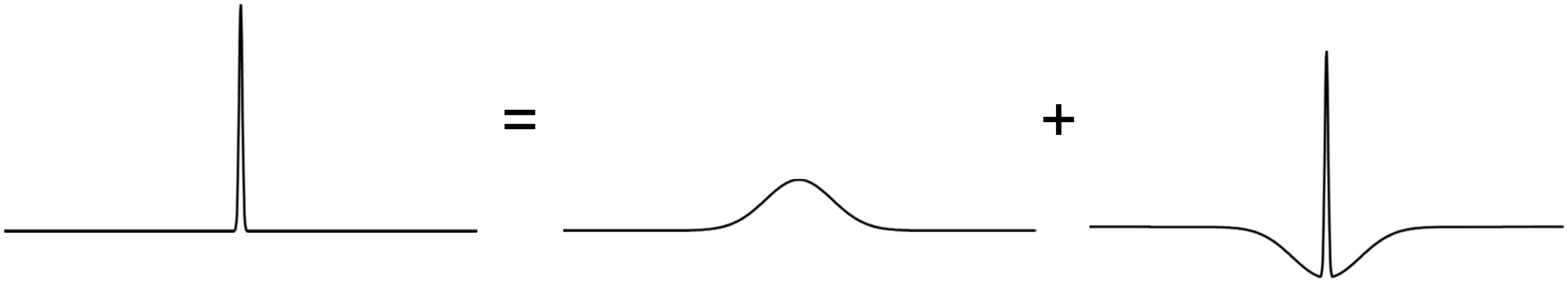
Electrostatic potential at the position of atom *i* is given by,

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

When **n** = 0, pairs of *i* = *j* are excluded.

Particle Mesh Ewald method (1)

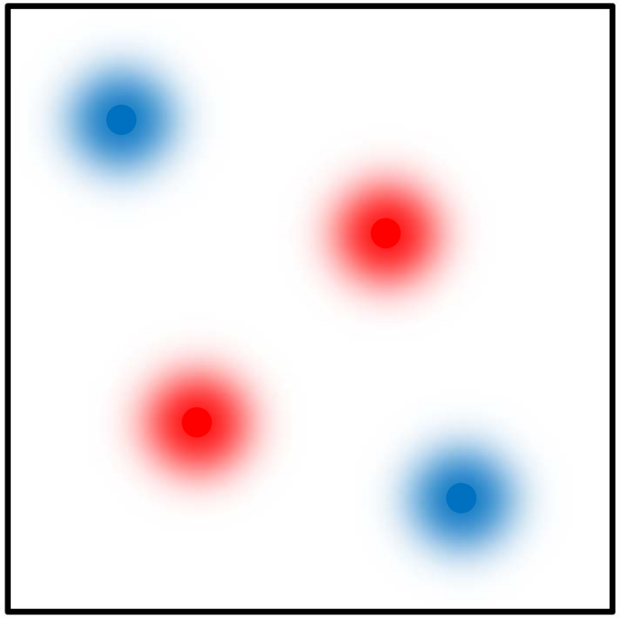
- Express the charge distribution of a point charge by the sum of two distributions



Point charge = Gaussian distribution + Residual distribution

$$\rho^{\sigma}(\mathbf{r}) = \sum_i q_i \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{3}{2}} \exp\left(-\frac{|\mathbf{r} - \mathbf{r}_i|^2}{2\sigma^2} \right)$$

Particle Mesh Ewald method (2)



- Gaussian distribution is smooth.
  →Electrostatic potential is calculated by solving Poisson’s equation with the fast Fourier transform.

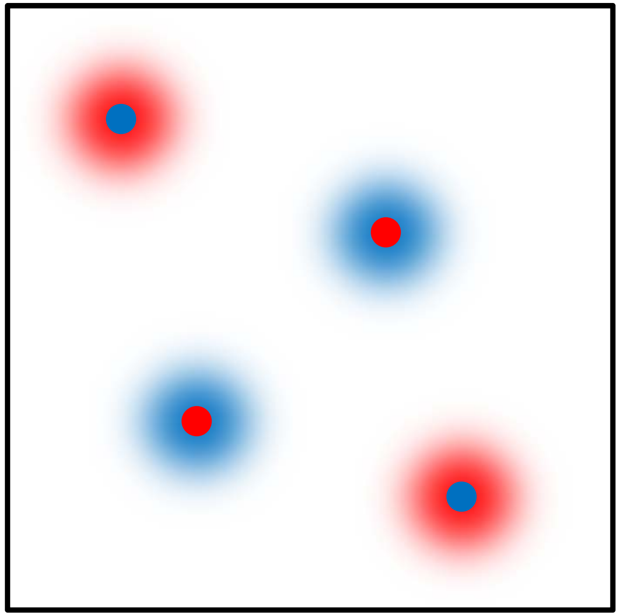
  $$\nabla^2 \varphi(\mathbf{r}) = -4\pi \rho^{\sigma}(\mathbf{r})$$

  $$\Big\Downarrow$$ Fast Fourier transform

  $$\tilde{\varphi}(\mathbf{k}) = \frac{4\pi}{|\mathbf{k}|^2} \tilde{\rho}^{\sigma}(\mathbf{k})$$

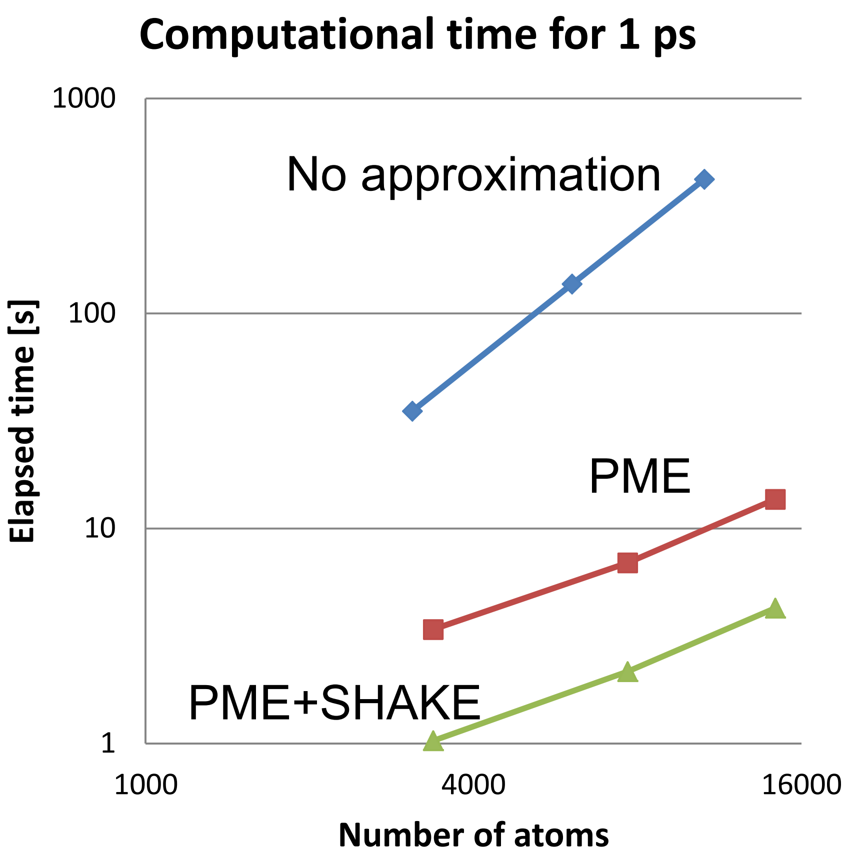
- To avoid divergence, total charges must be zero.

Particle Mesh Ewald method (3)



- 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.

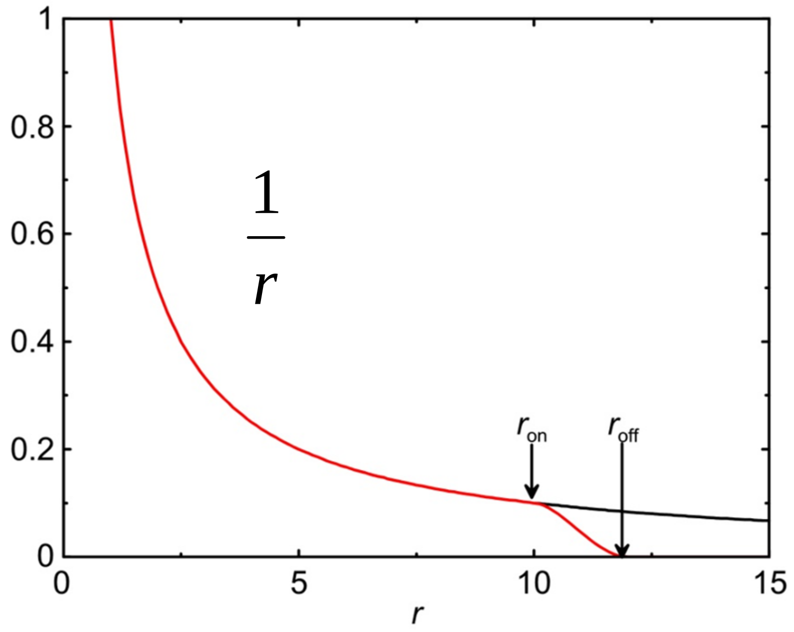
Computational time (2)

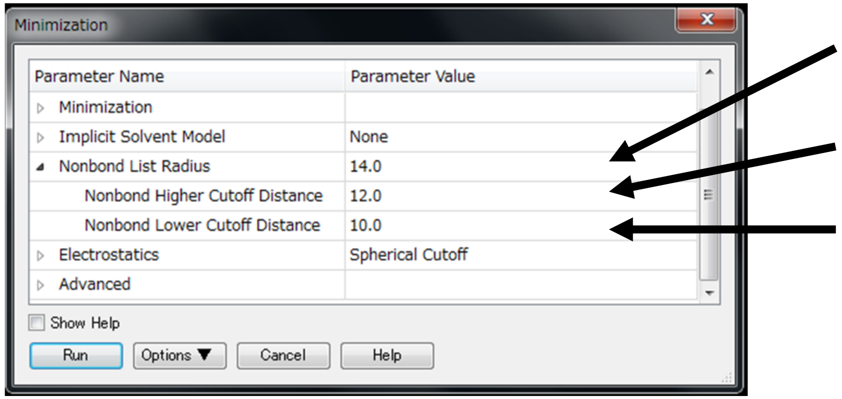


- Simulations were performed on water spheres.
- Without approximation, computational time is proportional to N^2.
- With PME, it becomes proportional to $N\log N$.
- SHAKE allows the use of four-times longer time step (2 fs). The computational speed becomes 3.2-times faster.

Settings in Discovery Studio (3)

- The cut-off method is used by default.
- A switching function is used to alleviate the instability of the energy caused by the cut-off operation.





Parameter Name	Parameter Value
Minimization	
Implicit Solvent Model	None
Nonbond List Radius	14.0
Nonbond Higher Cutoff Distance	12.0
Nonbond Lower Cutoff Distance	10.0
Electrostatics	Spherical Cutoff
Advanced	

☐ Show Help

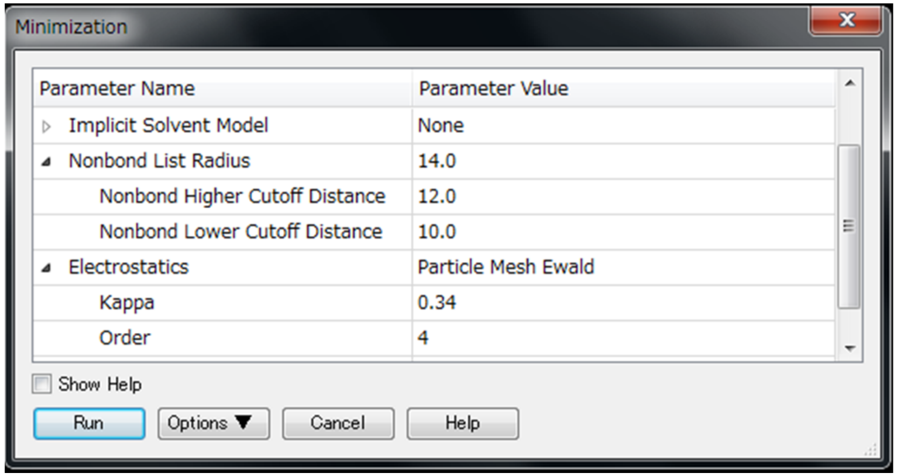
r_l (used in pair list generation)

r_{off}

r_{on}

Settings in Discovery Studio (4)

- Set Electrostatics to “Particle Mesh Ewald” to use this method.
- Van der Waals interactions are calculated with the cut-off method.
- Parameter σ is set via Kappa ($\kappa = 1/\sqrt{2}\sigma$). It should be between $3/r_{\mathrm{off}}$ and $6/r_{\mathrm{off}}$.
- If the system is not electrically neutral, add counterions at the solvation stage.



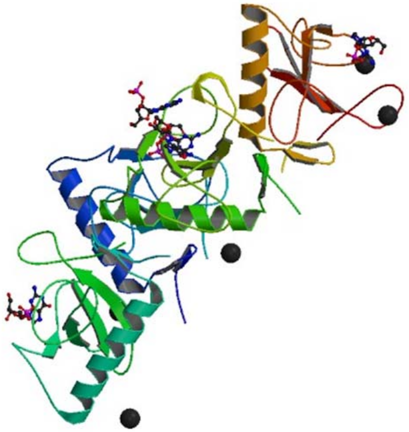
Parameter Name	Parameter Value
Implicit Solvent Model	None
Nonbond List Radius	14.0
Nonbond Higher Cutoff Distance	12.0
Nonbond Lower Cutoff Distance	10.0
Electrostatics	Particle Mesh Ewald
Kappa	0.34
Order	4

Procedures for MD simulation

1. Preparation of the initial structure
    - Obtain the structure
    - Add missing atoms and residues
    - Add hydrogen atoms
    - Obtain ligand force field parameters
    - Solvate the system
2. Energy minimization
3. Assignment of the initial velovities
4. Equilibration
5. Production

Preparation of the initial structure (1)

- Obtain the structure
  - – Download the experimental structure from PDB (http://www.rcsb.org/pdb/)
  - – Usually, simulations are performed for the biological units of the biomacromolecules.
  - – Example: Ribonuclease T1 (PDB ID: 1I0X)



Asymmetric unit

Biological Unit

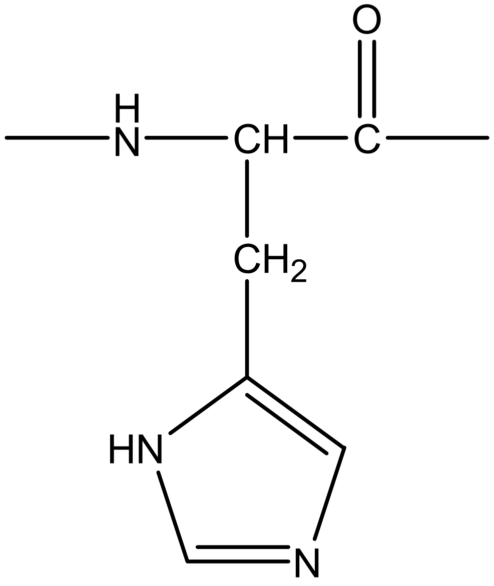
Preparation of the initial structure (2)

- Add missing atoms and residues
  - They can be added by using modeling software.
  - When N- or C-terminal residues are missing, you can block the terminus with an acetyl or N-methyl group.
- Add hydrogen atoms
  - Most of them are added automatically.
  - Pay special attention to SS bonds and protonation states of His.

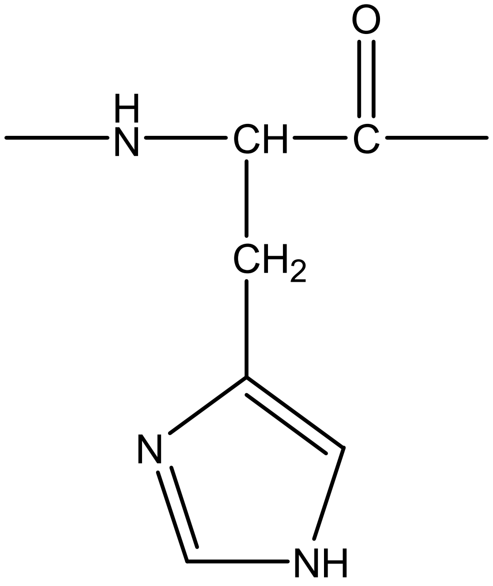
Operations in Discovery Studio (1)

1. Choose “File”→“Open URL” from the menu, enter “1I0X” for ID, and click “Open.”
2. Change Display Style to Line.
3. Select B, C, and D chains in Hierarchy Window and delete them.
4. Click “Macromolecules” button and expand “Protein Report” in the Tools tab.
5. Click “Protein Report.”
   →Check Incomplete or Invalid Residues. (Lys41, Asp49, Glu102 are colored purple.)
6. Expand “Prepare Protein” in the Tools tab, and click “Clean Protein” in the Manual Preparation section.
   →Missing atoms are added.

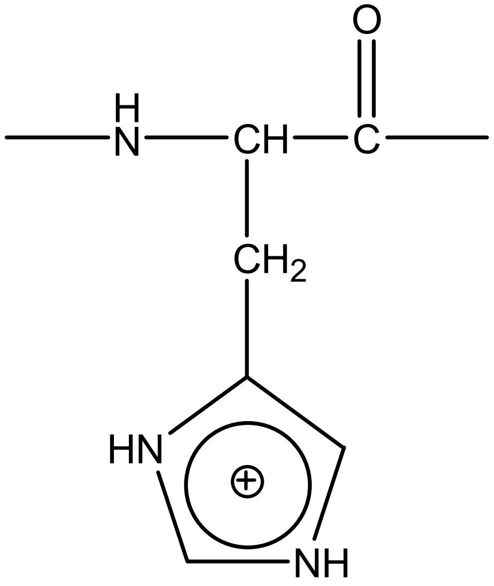
Protonation states of His



Protonation at δ



Protonation at ε

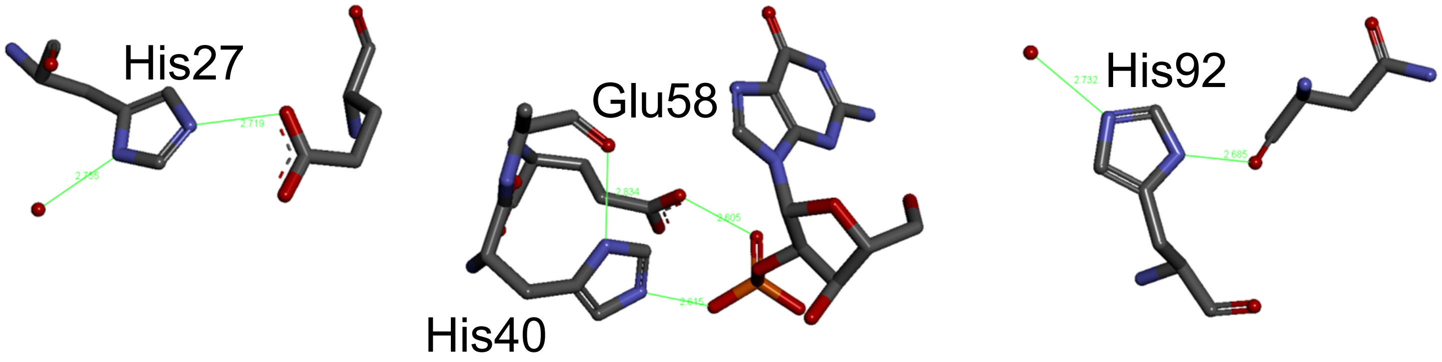


Protonation at δ and ε

- pK_a of His side chain is close to neutral (~ 6.5).
- You can find the protonation state from the hydrogen bond network where His is involved.

Operations in Discovery Studio (2)

7. Check the interactions of His27, His40, Glu58, and His92 with their surroundings.



8. Apply CHARMm force field, click “Calculate Protein Ionization and Residue pKa” in Protonate Protein section of Prepare Protein, and click “Run.”
→Check the protonation states of the residues.

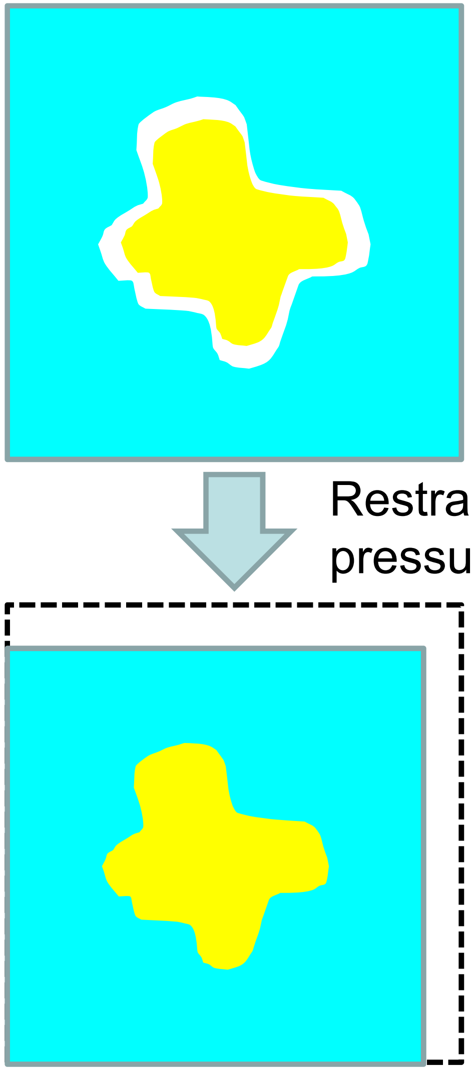
Preparation of the initial structure (3)

- Obtain ligand force field parameters
  - Ligand force field parameters are not included in the molecular dynamics software. It is necessary to make them by yourself or to obtain them from Amber Parameter Database.*
- Solvate the system
  - For an accurate and efficient simulation using the PME method, solvate system in a rectangular water box.
  - Add counterions to neutralize the system.

*http://www.pharmacy.manchester.ac.uk/bryce/amber

Equilibration

- In the initial structure, there is a space between the protein and the water.
- It is necessary to optimize water arrangement by performing a constant-pressure MD simulation.
- During the simulation, positions of protein atoms are restrained to their initial position and the restraints are gradually relaxed.



Space around the protein

Restrained constant-pressure MD simulation

Decrease of volume

How to send your report

- Use PowerPoint to create your report.
- Report should include the results and discussion of exercises 1 and 2.
- Send the PowerPoint file to tterada@iu.a.u-tokyo.ac.jp.
- Subject of the e-mail should be “Molecular modeling” and write your name and ID card number in the body of the e-mail.