今期から始まった遠隔授業の受講者の方々にも対応するため、早めにファイルを置いていますが、授業直前に内容を修正する可能性があります。2010年4月19日

#13
Molecular Simulation

Topics:

- Molecular simulation

  Quantum level, classical mechanics level
- Molecular Orbital Method (MO)

  Ab initio, semi-emprical, emprical, and DFT method
- Molecular Mechanics (MM)
- Molecular Dynamics (MD)
- Monte Carlo method (MC)

Molecular Simulation

- Quantum level (molecular orbital method)
  - Based on Schrödinger equation
  - Getting "electronic property" of the molecule

< 10,000 atom ?, for a snapshot

  - *ab initio method*

    Directly solving Schrödinger equation (Hartree-Fock approximation).

    Heavy requirement for calculation time and memory size. Best accuracy.
  - Semi-emprical method

    Using emprical value for electron repulsion integral, etc.

    Smaller calculation time and memory. Limited accuracy.
  - Emprical method

    Huckel approximation, etc.
  - Density Function Theory (DFT) method

    Accurate treatment for electron-electron interaction, with smaller cost than *ab initio* method.

Molecular Simulation

- Classical mechanics level

< 100,000 atom ?, < sub-micro sec. ?

  - Molecular Dynamics method
    - Based on Newton's motion equation for each atom in the molecule.
    - By integration of atomic force, time-series activity of the molecule can be investigated.
    - Cf. Integration with Molecular Orbital method

      First Principle Molecular Dynamics

      Car-Parrinello method (DFT+MD), QM/MM method

  - Monte Carlo method
    - By random sampling , numelically calculating the "partition function" of the phase space.
    - Static characteristic (free energy, etc.) are obtained.

MM: Molecular Mechanics

- Not a simulation. "Optimization" to a neighbor stable state.
  - Simulation = trying to mimic actual dynamical process.
  - Optimization = trying to quickly search stable state , by any mean.

- **Simple model for atomic energy**
  - bond length（結合長）energy： Hooke's spring model
  - bond angle （結合角）energy： Hooke's spring model
  - torsion angle （二面角）energy： cosine curve model for favored angles
  - uncharged atom interaction: Van der Waals interaction (L-J potential, etc.)
  - charged atom interaction: Coulomb interaction

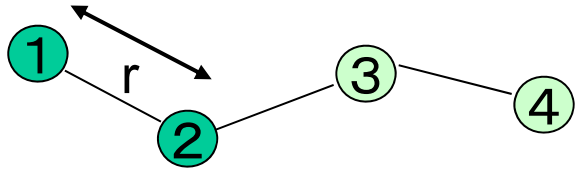
- **Step-wise movement toward lower-energy direction**
  - Steepest descent: Simply using first order differencial
  - Newton-Rapson method（Eigenvector Following）: Hessian (second order) matrix.
  - Broyden-Fletcher-Goldfarb-Shanno-Powell (F-P) : approximated inverse Hessian

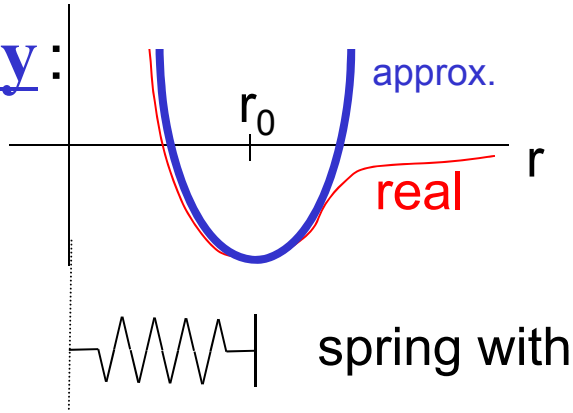
MM (bonded interactions)

– **bond length (結合長） energy**:

approximation by Hooke's law

Ebond = 1/2 * kb * $(r - r0)^2$





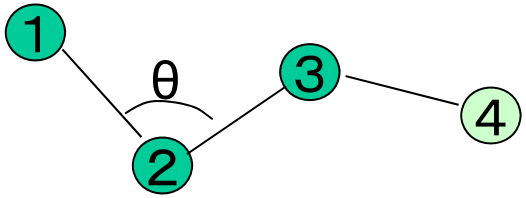
Potential

spring with length=r_0

cf) Morse function = $D \{1-\exp(-k(r-r_0))\}^2$

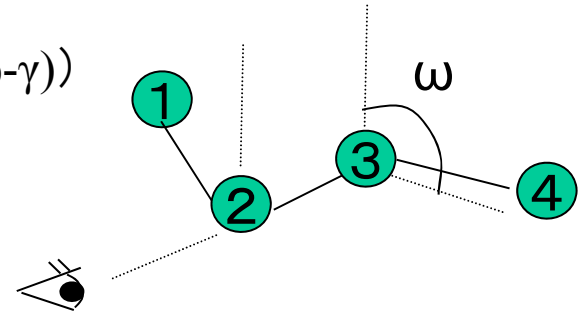
– **bond angle （結合角）energy**:

Eangle = 1/2 * ka * $(\theta - \theta 0)^2$



– **dihedral angle （二面角）energy**:

Etorsion = 1/2 * kt * $(1 + \cos(n\omega-\gamma))$

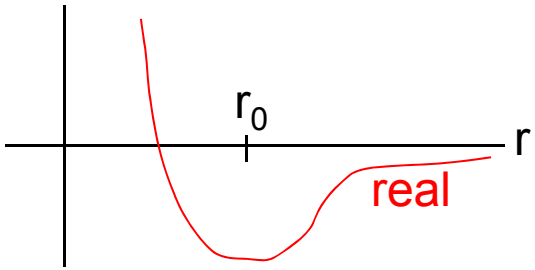


angle between
plane 1-2-3 and
plane 2-3-4

MM (non-bonded interactions)

– **uncharged atom energy** : Van der Waals interaction

Strong repulsion for close contact (r < r0) .
Weak attraction, even between uncharged atoms,
by induced dipole moment (r > r0).
Negligible interaction with remote atoms (r >> r0).



Hooke approximation is not suitable.

・rigid sphere (infinite repulsion for r < r0, otherwise zero)
・square well（r.s. ＋ weak attraction for an outer region)
・Sutherland （r.s. + weak attraction with long tail)
・Buckingham (repulsion:exponent, attraction: r^{-6})
・Lennard-Jones 6,12（repulsion: r^{-12}, attraction: r^{-6})

$$E_{LJ} = 4k\{(\sigma/r)^{12} - (\sigma/r)^{6}\}$$

Unique σ and k for each atom type.
σ is some Å. Cut-off approximation can apply.

Memo: bonded neighbor atoms should be omitted. 1-2, 1-3, and (1-4).

– **charged atom energy**: Coulomb interaction

$$E_{coulomb} = q1 * q2 / (4\pi\varepsilon * r)$$

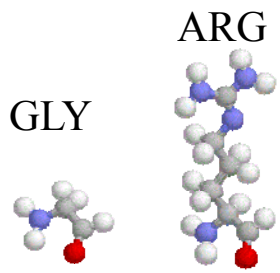
dielectric constant ε:
internal protein: ～1-10, solvent: ～ 80

Memo: Cut-off approximation of Coulomb force causes non-negligible error.

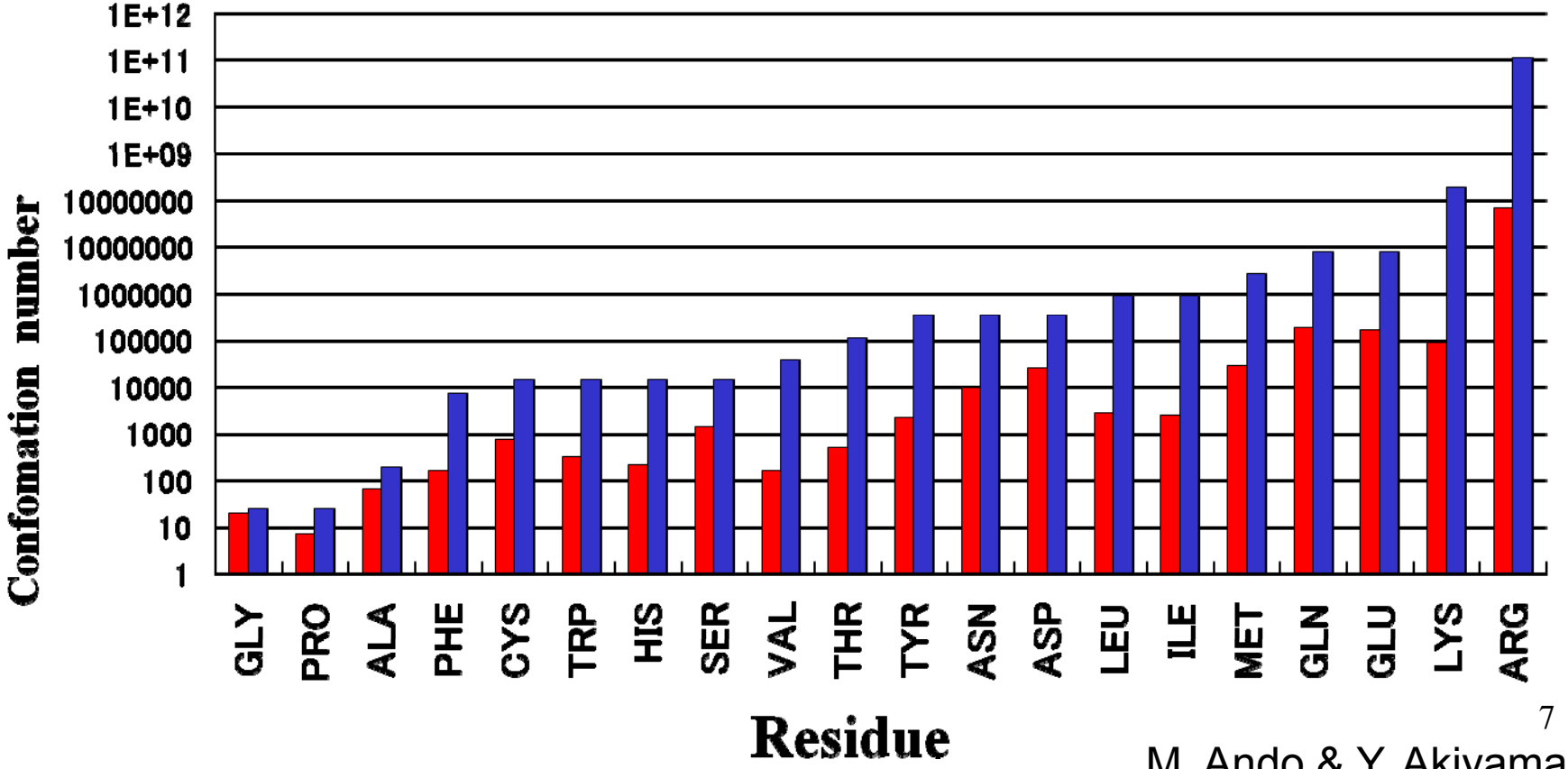
Huge conformational space..

Possible conformation of each residue (main+side chains)
(15 ° pitch)

GLY → 24
ARG → 110,075,314,176



Blue: simple multiplication of degree of freedom
Red: self-collision conformation excluded



M. Ando & Y. Akiyama

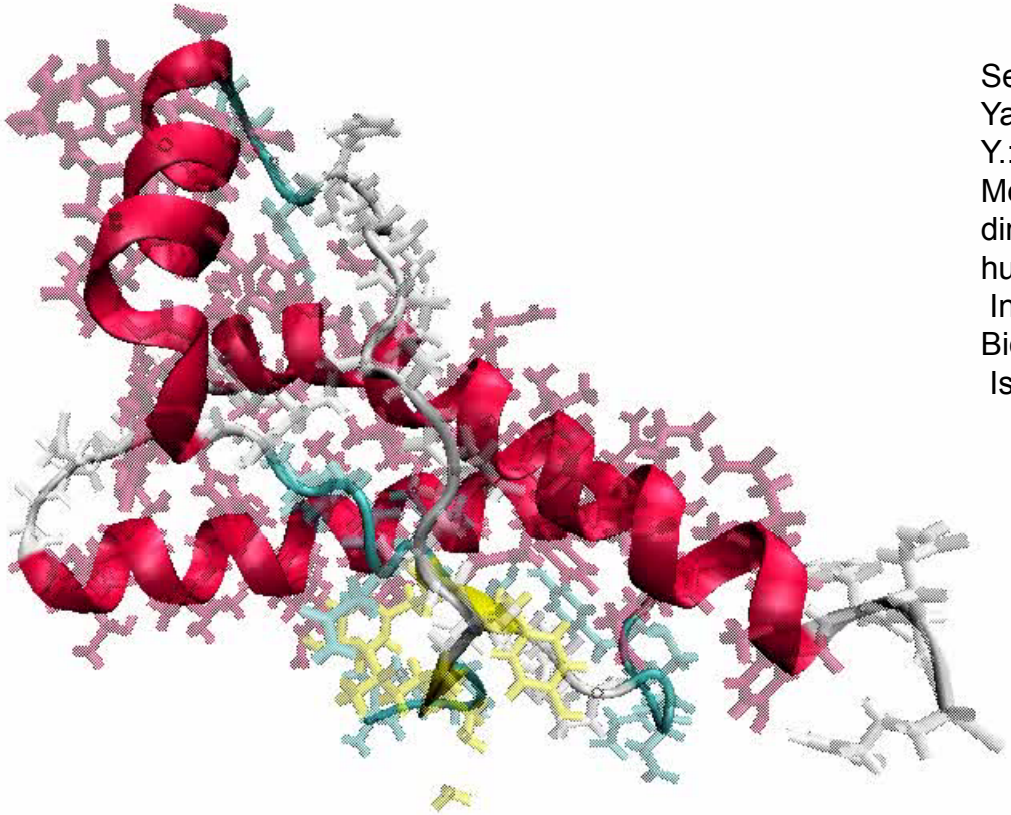
MD: Molecular Dynamics

- Simulating molecular particle motions with Newton's equation, by calculating atomic forces and its integration in time.

- **Calculation steps:**
  - 1. Set initial conditions (position, velocity)
    Starting from PDB coordinate, or starting from extended linear chain
  - 2. Calculate total force working on each atom
  - 3. Calculate acceleration. Update velocity and position
  - 4. (Check system temperature and pressure. Modify velocity, if needed)
  - 5. Go back to step2, unless enough simulation steps have been performed

- **Difference with Molecular Mechanics (MM)**
  - Each atom has its own "velocity". With high temperature, atoms have higher average velocity, and system can escape from local minimum status.
  - Purpose is different.

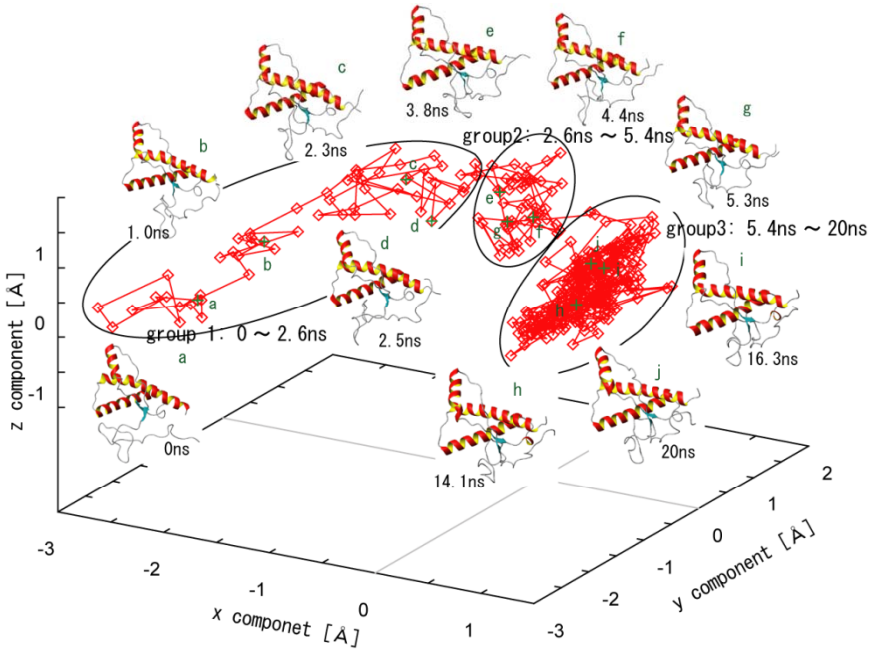
MD simulation of Prion protein



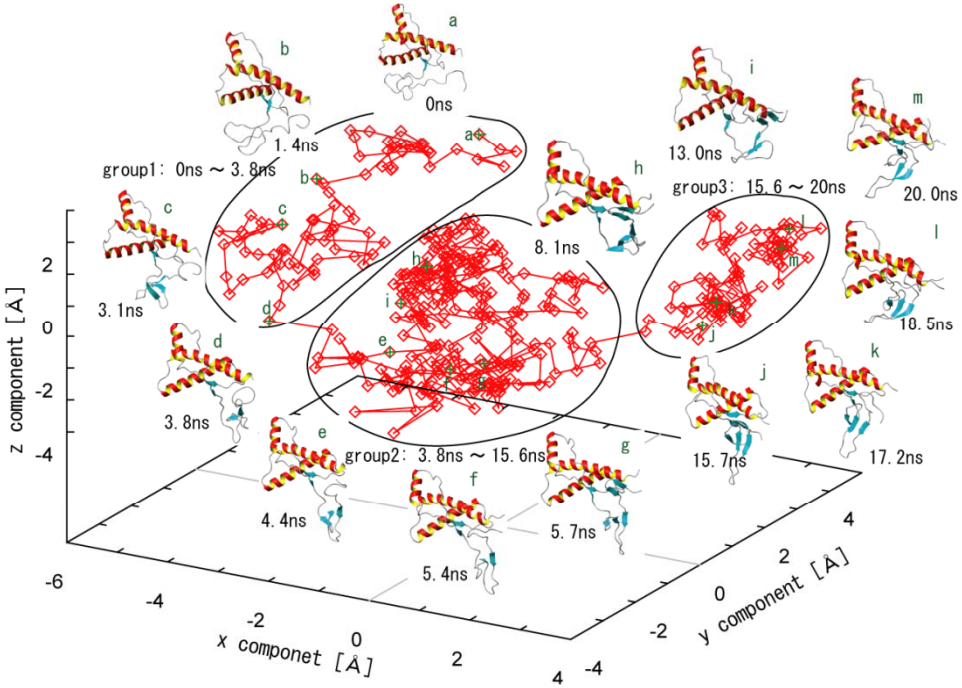
Sekijima, M., Motono, C., Yamasaki,S.Kaneko,K., and Akiyama Y.:
Molecular dynamics simulation of dimeric and monomeric forms of human prion protein (HuPrP): Insight into dynamics and properties, Biophysical Journal, Vol.85, Issue 2, pp.1176-1185, 2003.

Prion monomer
300K, 10ns
Purple: α helix
Yellow: β strand

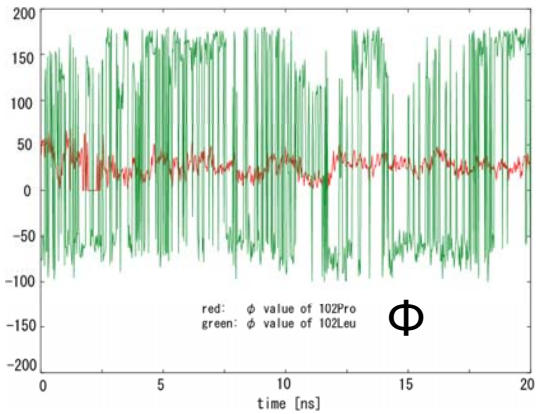
Influence of one-residue mutation

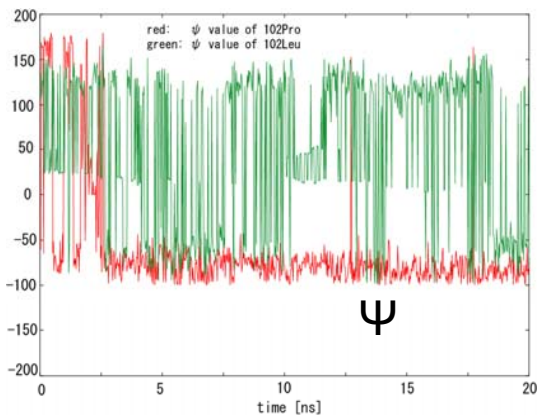


Native (102 Pro)



Mutation (102 Leu)

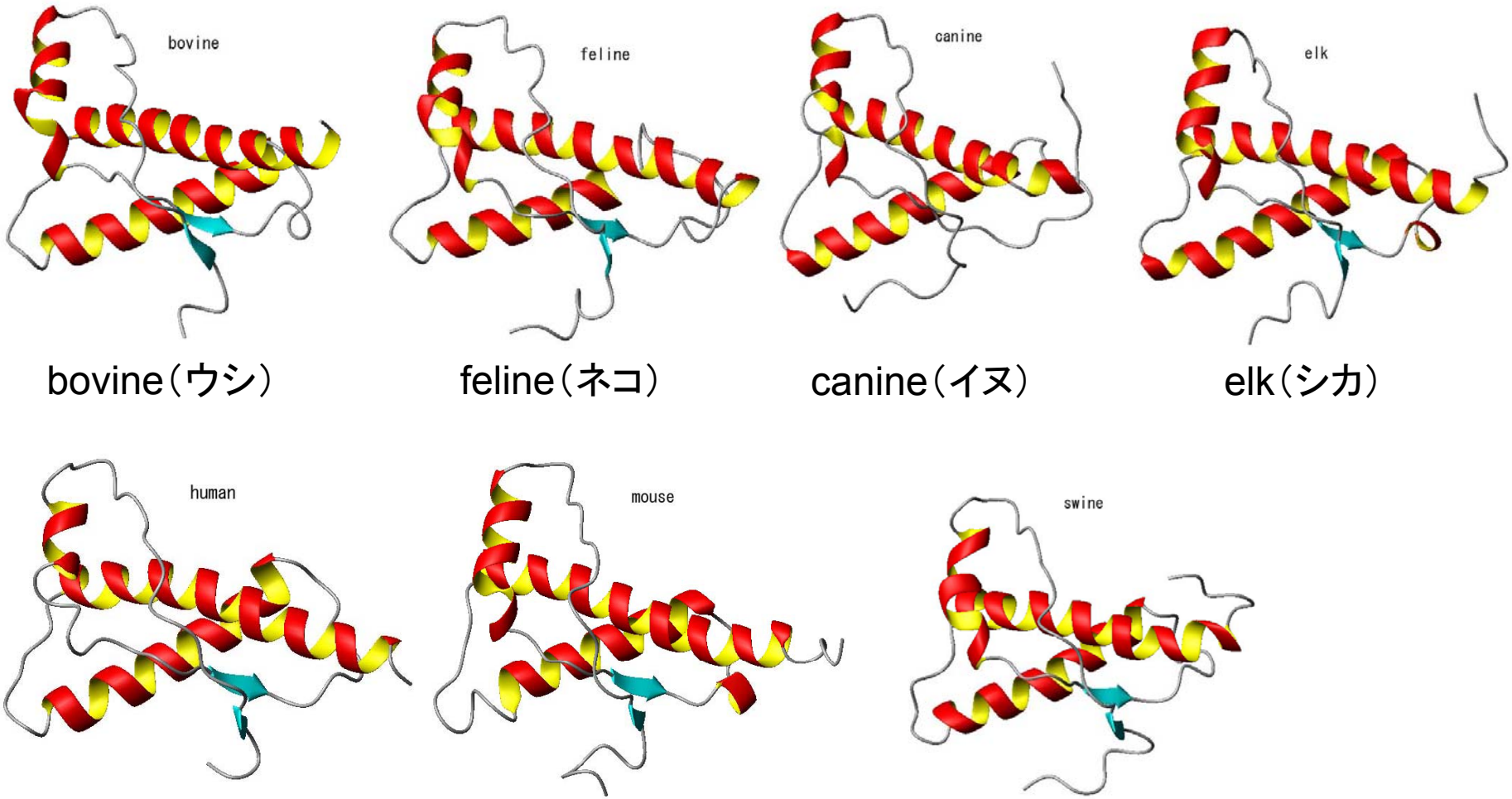




Pro (red) -> Leu (green)
Increased flexibility of
dihedral angle allowed
the change into β structure?

By courtesy of Prof. M. Sekijima (Tokyo Tech)

Comparison among different species



bovine（ウシ）　feline（ネコ）　canine（イヌ）　elk（シカ）

human（ヒト）　mouse（マウス）　swine(ブタ)

By courtesy of Prof. M. Sekijima (Tokyo Tech)

Major software for MD, MM

- AMBER (Kollman , UCSF)
  - http://amber.scripps.edu/
  - parm94, parm96, parm99 (parm94 tends to prefer α-herix , while parm96 does β-sheet)
- CHARMm (Karplus, Harvard univ.)
  - http://www.charmm.org/
- DISCOVER (Accelrys)
  - http://www.accelrys.com/insight/discover.html
- ECEPP (Sheraga, Cornell univ.)
  - http://www.tc.cornell.edu/CBIO/eceppak/
- GROMOS (Gunstern, ETH)
  - http://www.igc.ethz.ch/gromos/
- NAMD (Schulten, UIUC)
  - http://www.ks.uiuc.edu/Research/namd/
- OPLS (Jorgensen, Yale univ.)
  - http://zarbi.chem.yale.edu/

- Related topic 1： solvation

biomolecules work in solvation, not in vacuum.

  - **explicit water model**

    Solvation water is not merely H_2O molecule, it creates complex network.

    TIP3P, TIP4P（Jorgensen, Yale university） TIP4P assumes vertial charge

    ST2, SPC, etc.

  - **Implicit water model**

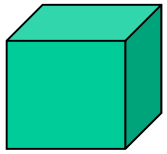
    Continuum body approximation（連続体近似）etc.

- Related topic 2： boundary

  - **Non-periodic boundary condition**

    Prepare large box or sphere, and fill capped water around proteins.

    Artificial force (pressure) are applied at boundaries.

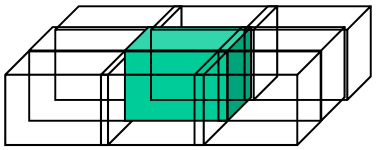


  - **Periodic boundary condition** （周期的境界条件）

    Assume infinite periodic copies of the basic region.

    No artificial force (pressure) at boundaries.

    （→Ewald method (FFT method) can be used.）



- Related topic ３: numerical integration
  - **Verlet method**
  - **Leap Frog method**
  - **RESPA**

- Related topic ４: calculation of $O(N^2)$ non-bonded interactions
  - **Distance cut-off approximation**

    Used traditionally, but several bad effects exist.

    Particles may suddenly “appear”/”disappear” at approximation radius.

  - **Tree method *O(N log N)*, FMM method *O(N)***

    use representative “cell” , instead of communicating with several particles.

    http://grape.astron.s.u-tokyo.ac.jp/~makino/papers/jnctam99_preprint/jnctam99_preprint.html

  - **Ewald method**

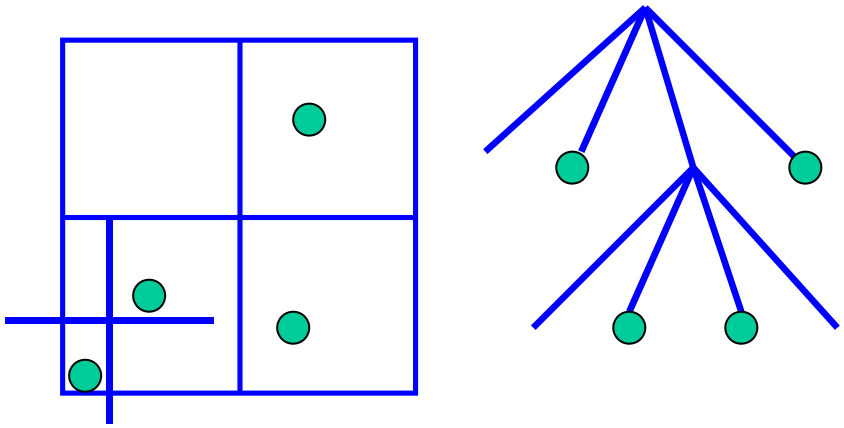
    direct calculation for local interactions.

    Fourier series expansion are used for remote interactions.

（advanced）

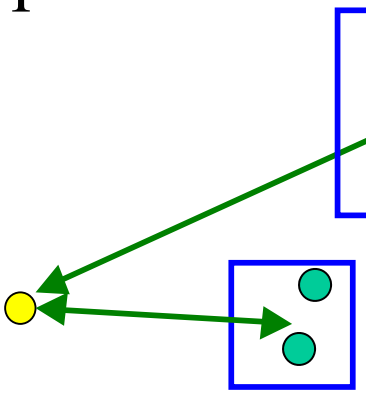
Barnes-Hut tree technique

- Space division by hierarchical cells.



2-D : quartet(4)-tree
3-D : octet(8)-tree

- Calculate interaction between a particle and a cell

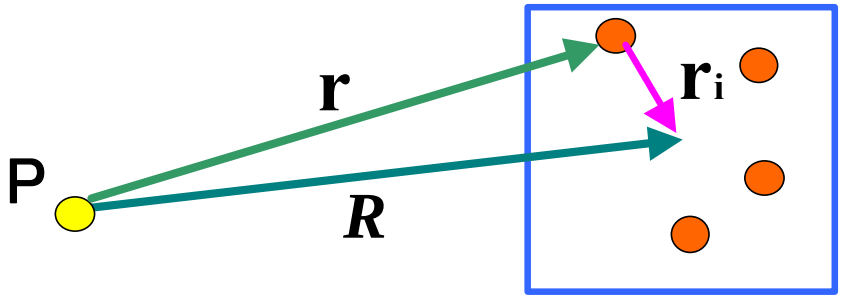


Opening angle (見込み角)

θ

Multi-pole expansion of charge distribution in a cell

（advanced）

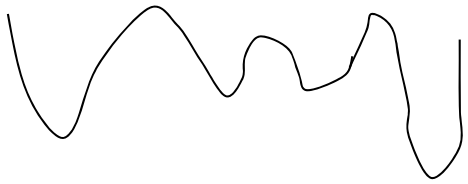


monopole dipole quadrupole

$$\Phi(R) = -\sum_i m_i \left[\frac{1}{R} + x_i \frac{x}{R^3} + y_i \frac{y}{R^3} + z_i \frac{z}{R^3} + \frac{1}{2} x_i^2 \left(-\frac{1}{R^3} + \frac{3x^2}{R^5} \right) + \frac{1}{2} y_i^2 \left(-\frac{1}{R^3} + \frac{3y^2}{R^5} \right) + \frac{1}{2} z_i^2 \left(-\frac{1}{R^3} + \frac{3z^2}{R^5} \right) + x_i y_i \left(\frac{3xy}{R^5} \right) + y_i z_i \left(\frac{3yz}{R^5} \right) + x_i z_i \left(\frac{3xz}{R^5} \right) \right]$$

MC: Monte Carlo method

- General technique for searching multi-dimensional problem space



- **Calculation steps**
  - 1. Set initial solution candidate X_0
  - 2. Generate new solution candidate X (usually, from neighbor of X_0)
  - 3. Compare cost (or energy) of two candidates X_0 and X.
    - **In case of using "Metropolis criterion" (メトロポリス基準)**
      - $E(X) < E(X_0)$, always accept improved new X.
      - $E(X) >= E(X_0)$, sometime accept non-improving X, with an accept probability of

        $P = \exp (- (E(X) - E(X_0)) / T)$ T: temperature parameter

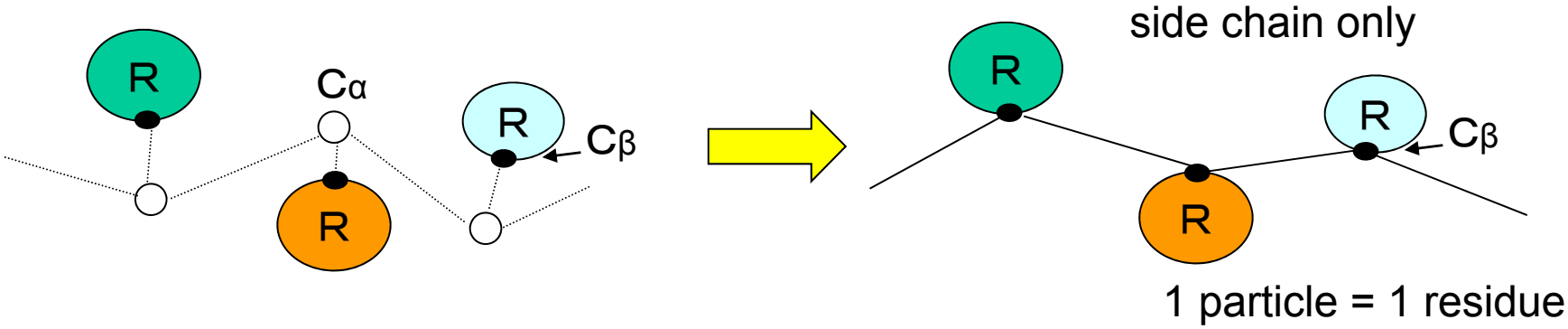
        If new X is accepted, set new $X_0 = X$
  - 4. Stop, if current solution candidate X_0 is favorable, otherwise go back to Step 2.

- Characteristics
  - With higher temperature T, various status are sampled, escaping from local minima.
  - Generation of new good candidate (from neighbors) is a key issue.
  - Bad performance like a random walk, if candidate generation is not efficient.

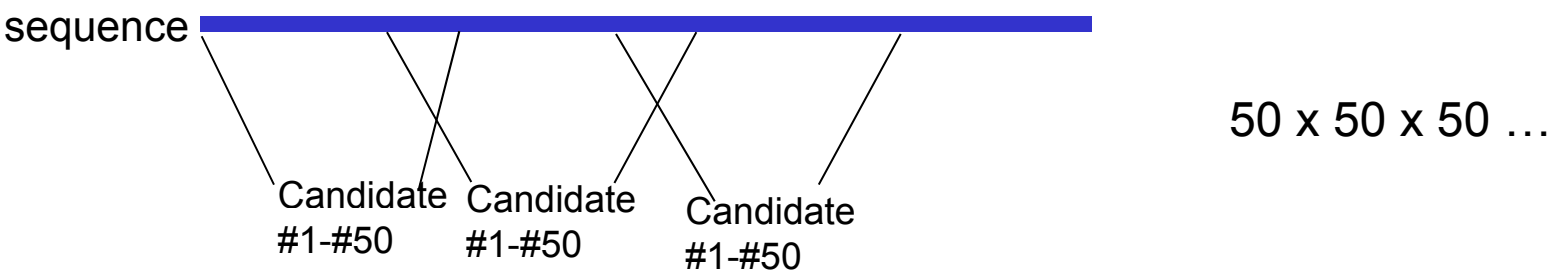
Protein Structure Prediction by MC

Coarse models are used.

- Jeff. Skolnick (Buffero Univ.): SICHO model



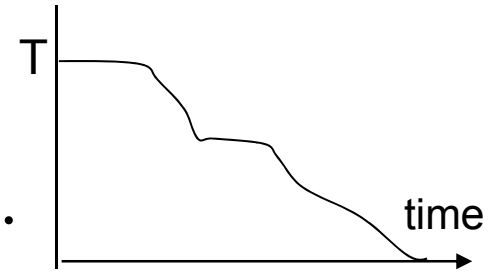
- David Baker (Washington U.): Fragment Assembly
  - http://depts.washington.edu/bakerpg/



3-D Lattice model -> off-lattice Monte Carlo sampling

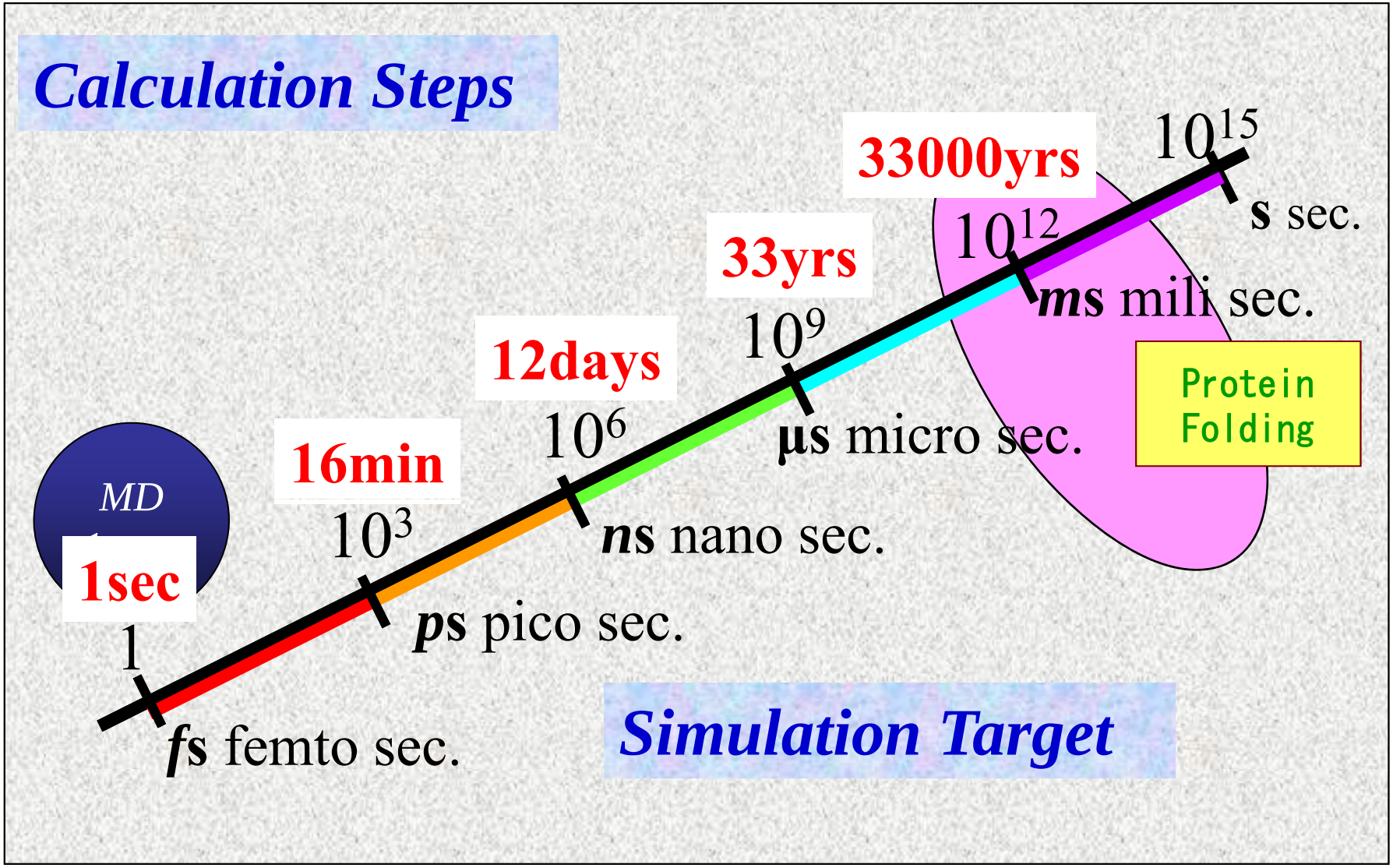
Solution Techniques

- Simulated Annealing
  - Start from High T, and slowly decrease.
  - Combination with MC, and MD
  - Cf.) Specific heat（比熱）dependent scheduling

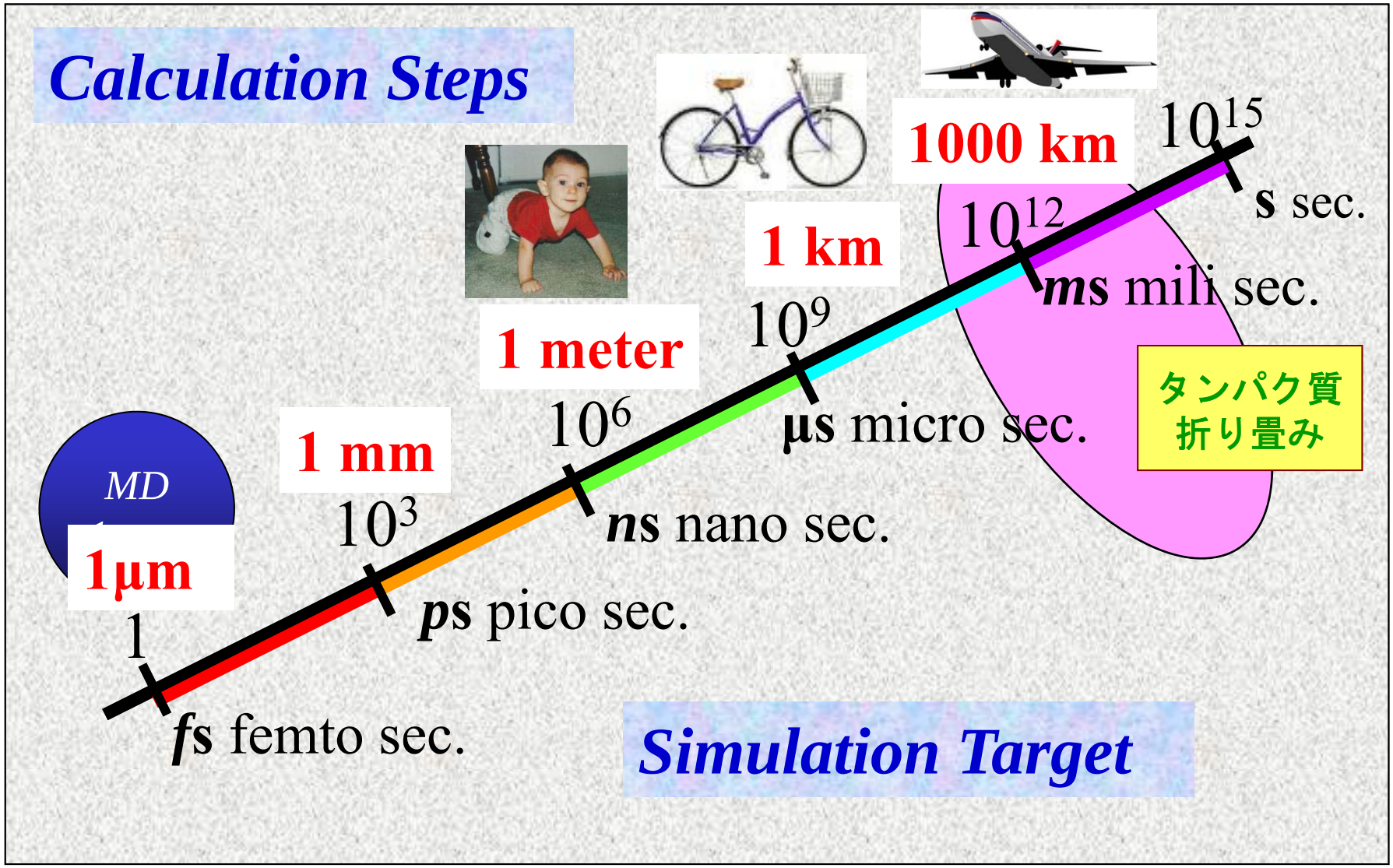


- Replica Exchange method
  - Prepare multiple fixed-temperature "baths" $\{T_1, T_2, T_3, \ldots T_n\}$.
  - Perform single MC search in each bath T_i.
  - Occasionally exchange solutions between neighboring "baths" T_i and T_{i+1}, based on Metropolis criterion.
  - Each MC search experiences several different temperatures.

Calculation Time for MD Simulation

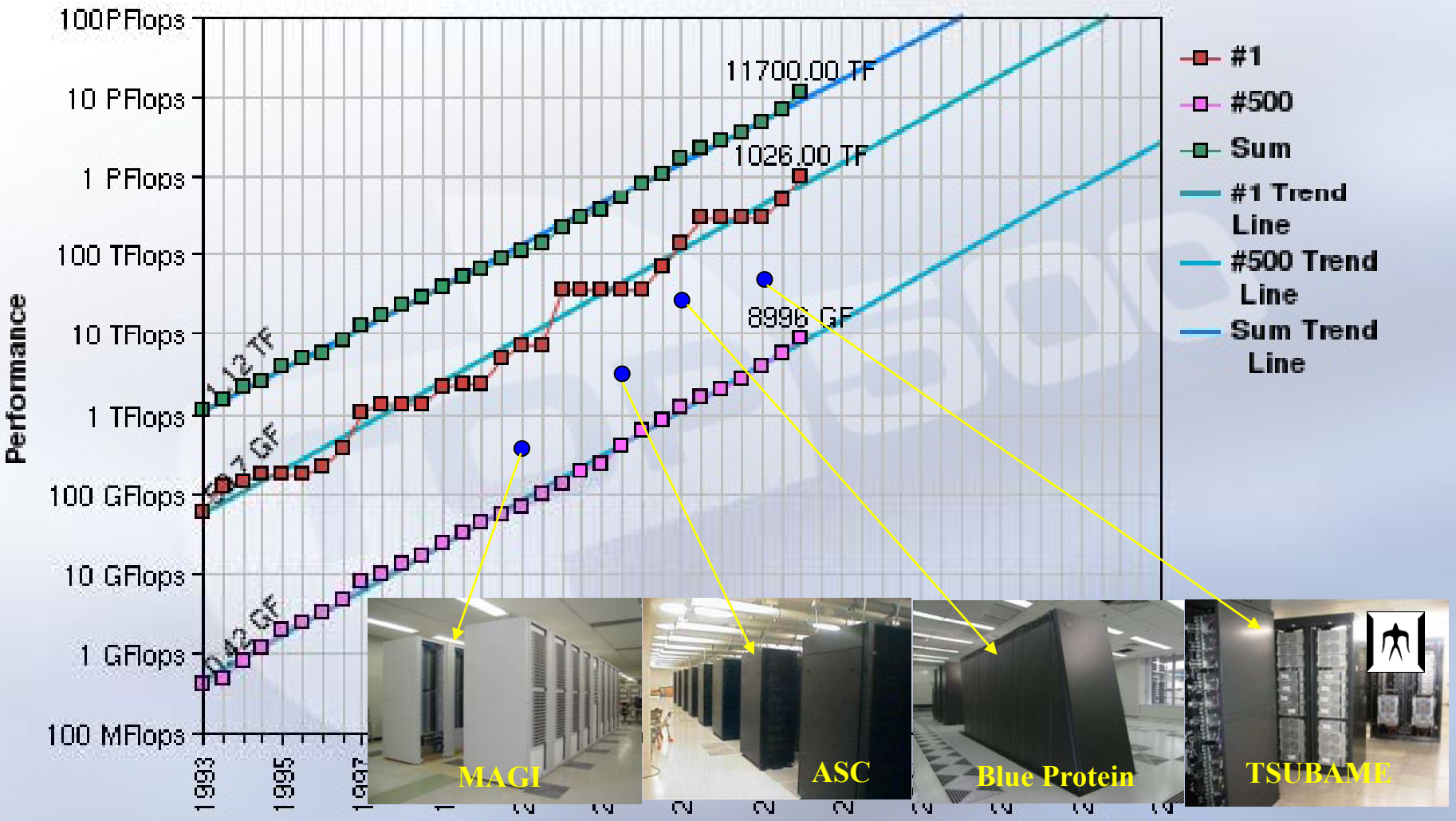


Calculation Time for MD Simulation



Progress of Computer Performance

From www.top500.org



1000-fold growth / 10-12 yrs

東京工業大学 学術国際情報センター TSUBAME Grid Cluster

大規模分散並列型演算サーバ　高性能密結合分散並列型演算サーバ　メモリ共有型演算サーバ群

Sun Fire X4600
655nodes
16CPU/node
10480CPU
Memory: **21.4TB**

type	nodes	CPU clock	MEM/node
A	639	2.4GHz	32GB
B	16	2.6GHz	64GB

分野	アプリケーション名
計算化学	Gaussian03
	GaussView
	TCP Linda for Gaussian
	Molpro
	MOPAC
	AMBER 8
	Materials Explorer
	Materials Studio
	Discovery Studio Modeling
構造解析	ABAQUS (Standard,Explicit)
	ABAQUS/CAE
	NASTRAN
	PATRAN
統計解析	SAS
	MATLAB
数値計算数式処理	Mathematica
可視化	AVS/Express PCE
	AVS/Express Developer
	EnSight

ClearSpeed 360枚

・・・655nodes

InfiniBand Network Voltaire ISR 9288 ×8

100ギガビット級ネットワーク装置

24Gbps
(片方向)

SuperTitanet

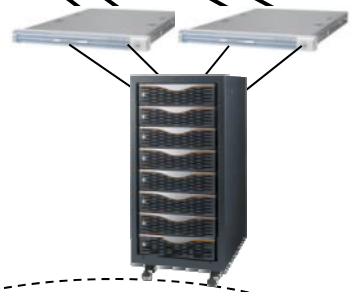
200bps
(片方向)

・・・
42台

500GB
48disks

500GB
48disks

500GB
48disks



ストレッジサーバA
Sun Thumper (Code Name)
物理容量 1PB

ストレッジサーバB
NEC iStorage S1800AT
物理容量 96TB RAID6

ペタバイト級ストレッジサーバ

48.88 TFLOPS
Linpack benchmark
(14th in world,
2007/06)
Upgraded
87.01 TFLOPS
(41th 2009/6)