



GENESIS

Generalized-ensemble simulation system

CBI学会2017年大会

GENESISチュートリアル(座学)

理化学研究所・生命システム研究センター
理化学研究所・計算科学研究機構

杉田有治

2017年10月5日

今日の内容

- 13:30 – 14:00 座学
 - 生体分子の分子動力学計算
 - GENESISの特徴
- 14:00 – 14:05 休憩・準備
- 14:05 – 15:30 実習1
 - GENESISのコンパイル、基本の利用方法
- 15:30 – 15:45 休憩
- 15:45 – 17:00 実習2
 - GENESISでの自由エネルギー計算

分子動力学法(Molecular Dynamics; MD)

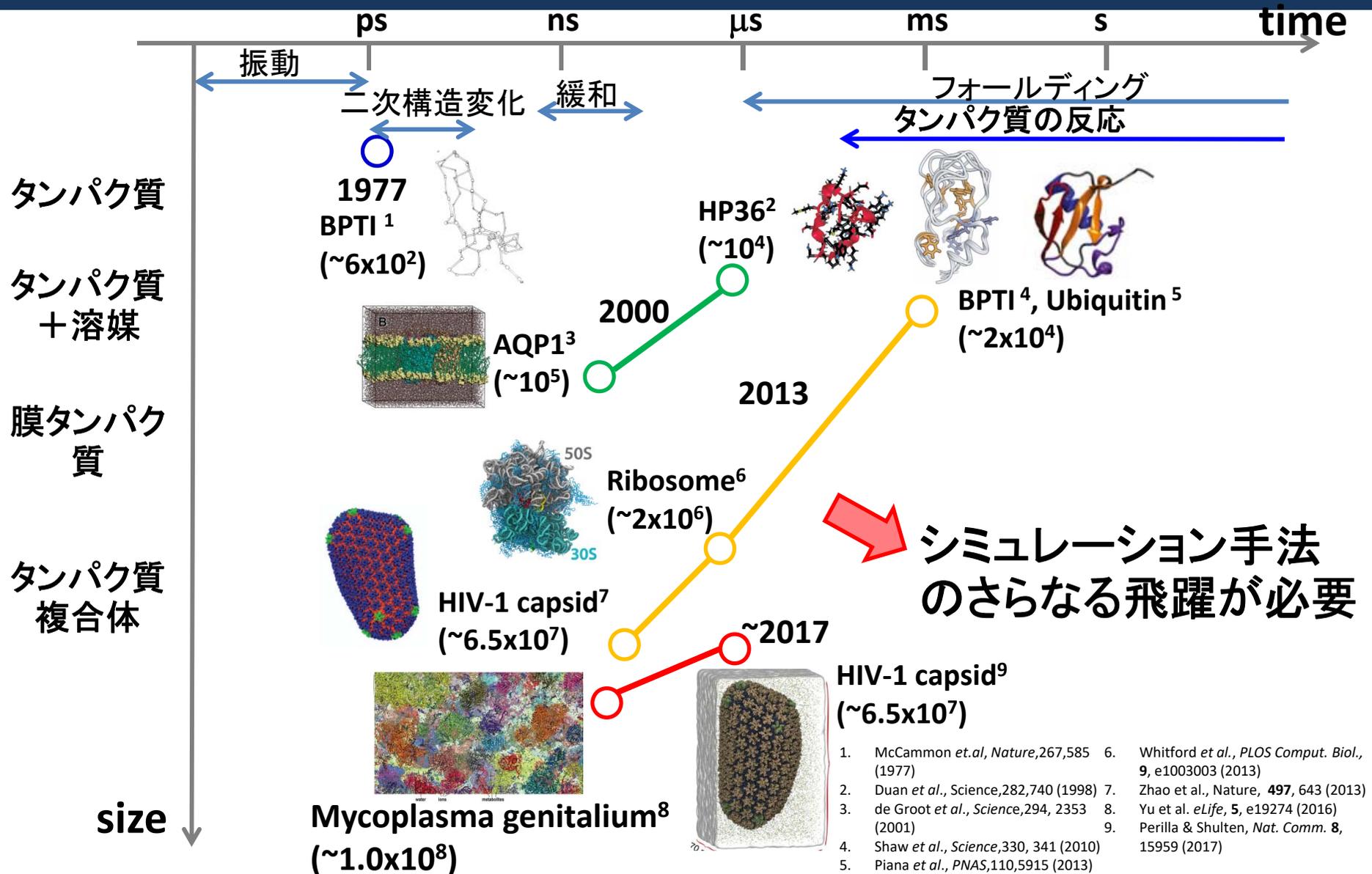
- 粒子間の相互作用力を計算し、ニュートンの運動方程式を解く

$$m_i \frac{d^2 \mathbf{r}_i(t)}{dt^2} = \mathbf{F}_i(t) = - \frac{dU(\mathbf{r}^N)}{d\mathbf{r}_i}$$

相互作用エネルギー ←
← 粒子の座標

- 粒子の運動から、構造の安定性、構造ゆらぎ、更にアンサンブルを計算し、自由エネルギー面(結合エネルギーなど)を解析できる
- しかし、MD計算はとても時間がかかる...
 - タンパク質のダイナミクスをミリ秒計算するには、ニュートンの運動方程式を 10^{12} 回計算する必要がある
 - 例えば1stepに1msかかるなら、約31.7年必要!

タンパク質のシミュレーション



1. McCammon *et al.*, *Nature*, 267, 585 (1977)
2. Duan *et al.*, *Science*, 282, 740 (1998)
3. de Groot *et al.*, *Science*, 294, 2353 (2001)
4. Shaw *et al.*, *Science*, 330, 341 (2010)
5. Piana *et al.*, *PNAS*, 110, 5915 (2013)
6. Whitford *et al.*, *PLOS Comput. Biol.*, 9, e1003003 (2013)
7. Zhao *et al.*, *Nature*, 497, 643 (2013)
8. Yu *et al.*, *eLife*, 5, e19274 (2016)
9. Perilla & Shulten, *Nat. Comm.* 8, 15959 (2017)

重点課題1: 創薬計算の高精度化・高速化

- 基質結合とタンパク質構造変化の両方を取り込む
- スパコンを活用した高並列スクリーニング

創薬計算に主に用いられているMDソフトと力場

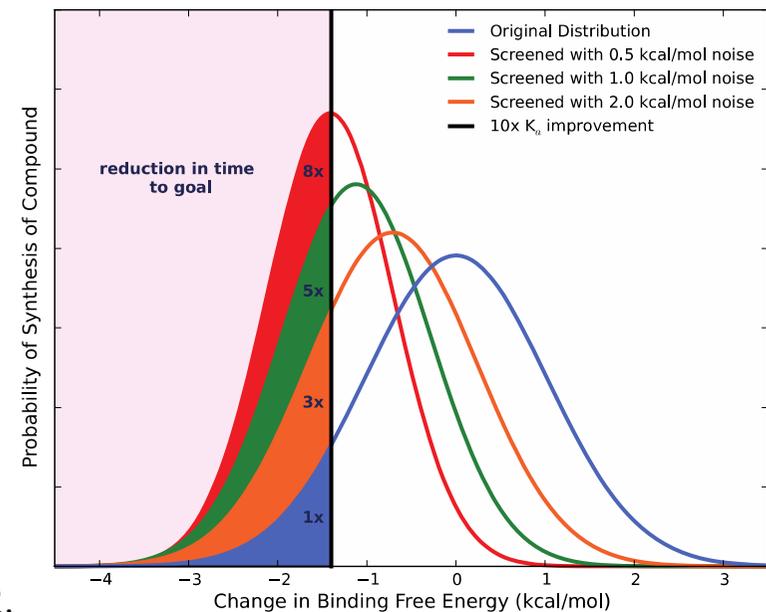
TABLE 1

Main software and force-fields used in MD.

Developing group	MD suite	Force field
Groningen (NL)	GROMACS [1]	GROMOS [16]
Harvard (USA)	CHARMM [3]	CHARMM [15]
San Francisco (USA)	AMBER [4]	AMBER [13]
Zurich (CH)	GROMOS [22]	GROMOS [16]
Purdue/Yale University (USA)	-	OPLS [14]
Urbana-Champaign (USA)	NAMD [2]	-
Barcelona (ES)	ACEMD [23]	-
New York (USA)	DESMOND	-

(Mortier, J. et al. *Drug Discovery Today*, **2015**, 20:686-702,
Mobley, D. & Klimovich, P. V. *J. Chem. Phys.* **2012**, 137:230901.)

コンピュータスクリーニングの導入で
合成効率はどう変わる？





GENESIS

Generalized-ensemble simulation system

- 生体系において、効率的で精度の良い自由エネルギー計算手法の開発が目的
- 高並列計算 - 京コンピュータなどでの超並列計算が可能
(ただし、普通のPCクラスタでも動きます)
- 巨大な生体系も可能
- 全原子モデルのみでなく、粗視化モデル等異なる分子モデルへも応用できるアルゴリズムを採用
- レプリカ交換法(拡張アンサンブル法の一つ)による自由エネルギー計算も可能



GENESIS

Generalized-ensemble simulation system



■ 開発チーム

計算科学研究機構(AICS)・粒子系生物物理研究チーム

プロジェクトリーダー: 杉田有治

主開発者: 鄭載運、森貴治、小林千草、松永康佑、神谷基司、
田村康一、優乙石、八木清

公開バージョン : Ver. 1.1.5

■ ホームページ

<http://www.aics.riken.jp/labs/cbrt/>

ソースコードとマニュアルは上記ページ内のリンクよりダウンロード可

■ 参考文献

J. Jung, T. Mori, C. Kobayashi, Y. Matsunaga, T. Yoda, M. Feig, and Y. Sugita, WIREs Comput. Mol. Sci., **5**, 310-323 (2015).

C. Kobayashi, J. Jung, Y. Matsunaga, T. Mori, T. Ando, K. Tamura, M. Kamiya and Y. Sugita, *J. Comput. Chem.* **38**, 2193-2206 (2017).

生体分子MD計算のボトルネック

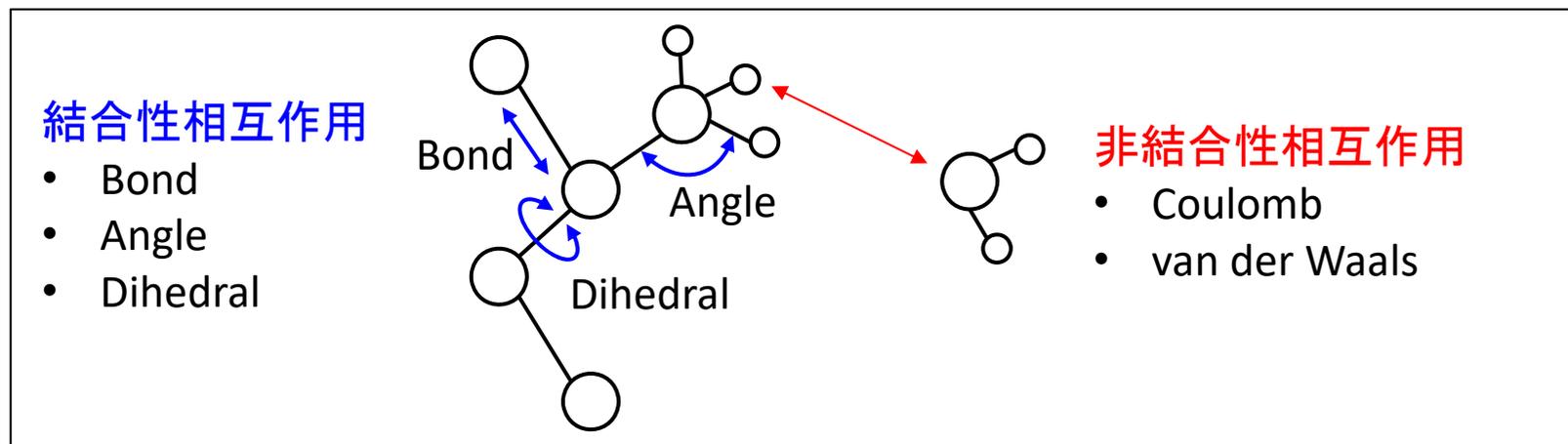
- 分子動力学法は膨大な回数のニュートン方程式を解く
- **最も時間がかかる部分は粒子間の相互作用計算の部分である**

生体分子の相互作用計算は「力場(force field)」と呼ばれる経験的な関数で記述される

force fieldはおおまかに2つの部分に分けられる

- **結合性相互作用**: 原子間の共有結合によるもの
- **非結合性相互作用**: 長距離相互作用(電荷、van der Waals力)によるもの

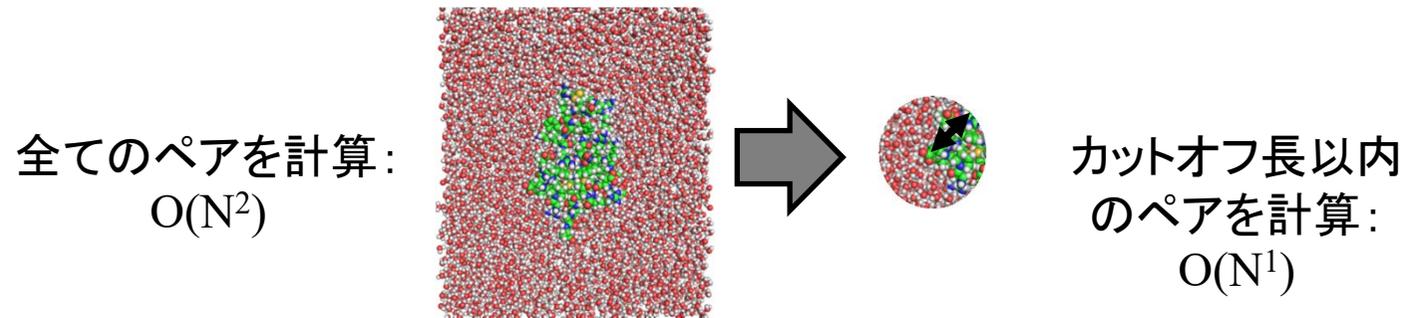
粒子数(N)に対して $O(N^2)$ で時間がかかるため、最も時間のかかる部分



→ 非結合相互作用の高速化が必要

非結合相互作用の高速化

1. カットオフ長を導入することで $O(N^2)$ から $O(N^1)$ に時間短縮



2. カットオフ長より長い部分のクーロン相互作用はフーリエ変換により逆空間で計算

$$U_{elec} = \underbrace{\sum_{|i-j| < R} \frac{q_i q_j}{4\pi\epsilon_0} \frac{\text{erfc}(\alpha_{ij})}{r_{ij}}}_{\text{実空間(短距離)}} + \underbrace{\frac{2\pi}{V} \sum_{\mathbf{G} \neq 0} \frac{\exp(-|\mathbf{G}|^2 / 4\alpha^2)}{|\mathbf{G}|^2} \sum_{ij} \frac{q_i q_j}{4\pi\epsilon_0} \cos(\mathbf{G} \cdot \mathbf{r}_{ij})}_{\text{逆空間(長距離)}} - \underbrace{\sum_i \frac{q_i q_i}{4\pi\epsilon_0} \frac{\alpha}{\sqrt{\pi}}}_{\text{Self energy(定数)}}$$

3. 1,2に加えて、非結合相互作用の高速化、高並列化は短距離項、長距離項の双方で更に行われる

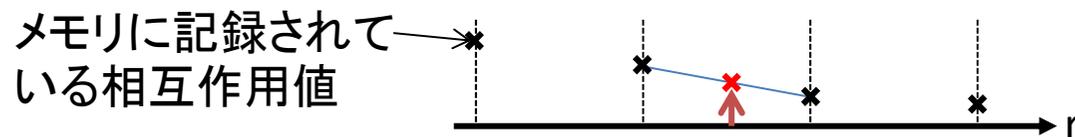
GENESISの高速化・高並列化

- 最も時間のかかる非結合相互作用計算を高速化するため
GENESISでは下のアルゴリズムを新規に開発、組み込む
公開版で導入された高速化
 - ✓ **Inverse lookup table approach** (短距離相互作用計算の高速化)
(Jung *et al.*, *J. Comput. Chem.*, **34**:2412–2420 (2013))
 - ✓ **Midpoint cell methods** (短・長距離相互作用計算の高並列化)
(Jung *et al.*, *J. Comput. Chem.*, **35**:1064–1072 (2014))
 - ✓ **Parallelization of FFT** (長距離相互作用計算の高並列化)
(Jung *et al.*, *Comp. Phys. Comm.*, **200**: 57-65 (2016))
 - ✓ **Usage of GPGPU** (GPGPUを用いた短距離相互作用計算の高速化)
(Jung *et al.*, *J. Chem. Theory Comput.* **12**:4947-4958 (2016))

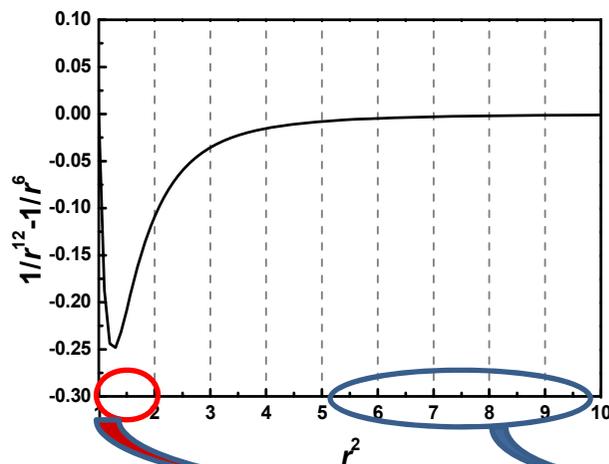
Inverse lookup table法

Jung et al., *J. Comput. Chem.*, **34**:2412–2420, (2013)

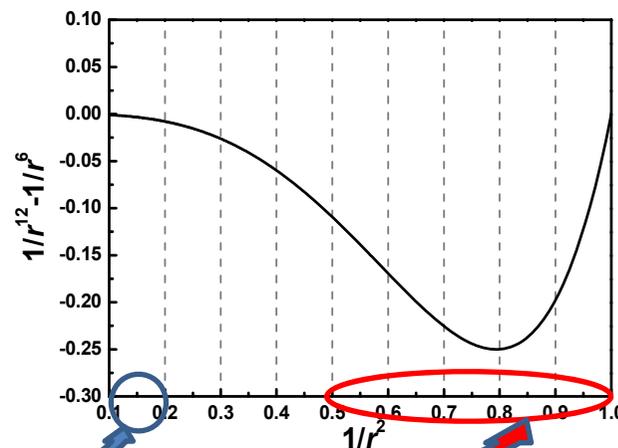
- 短距離相互作用は距離の関数であるため、カットオフ長までの距離の代表点での相互作用の値を計算、メモリ(table)に記憶しておく。



従来の方法



GENESISの方法



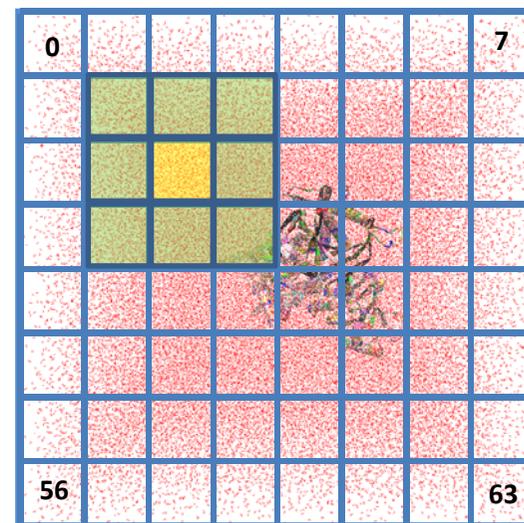
- 従来の方法では r^2 の線形関数、3次関数で内挿していた物を、GENESISでは $1/r^2$ の線形関数で内挿し、高速で精度の良い計算を可能とする

Midpoint cell 法

Jung et al., *J. Comput. Chem.*, 35:1064–1072, (2014)

■ ドメイン分割法 (ほとんどのMDプログラムで採用)

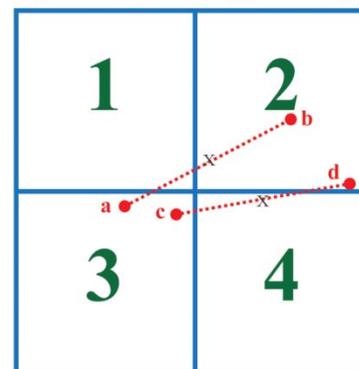
- 系全体をカットオフ長より長い一辺をもつセルで区切る
- エネルギー計算は隣接のboxのみを考えればよい (黄色のセル内の原子の相互作用は黄と緑のセル内の原子のみをカウントすればよい)
- 通信回数が減少されるため、並列度が良くなる
- ドメイン分割法で異なるセル間の相互作用計算をどのCPUコアに割り振るかが並列度に重要な問題になる



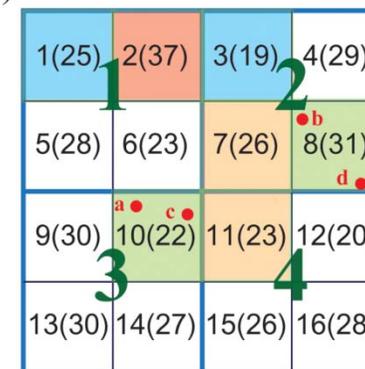
■ Midpoint cell法

- 従来のmidpoint法では、2つの原子の間地点のセルを受け持つコアが計算する
- それぞれの原子が存在する二つのセル (8, 10とする) の中間のセル (7, 11のどちらか) を受け持つコアが計算を担当する
- 通信回数が減少されるため、並列度が良くなる

(a) Midpoint法¹



(b) Midpoint cell法

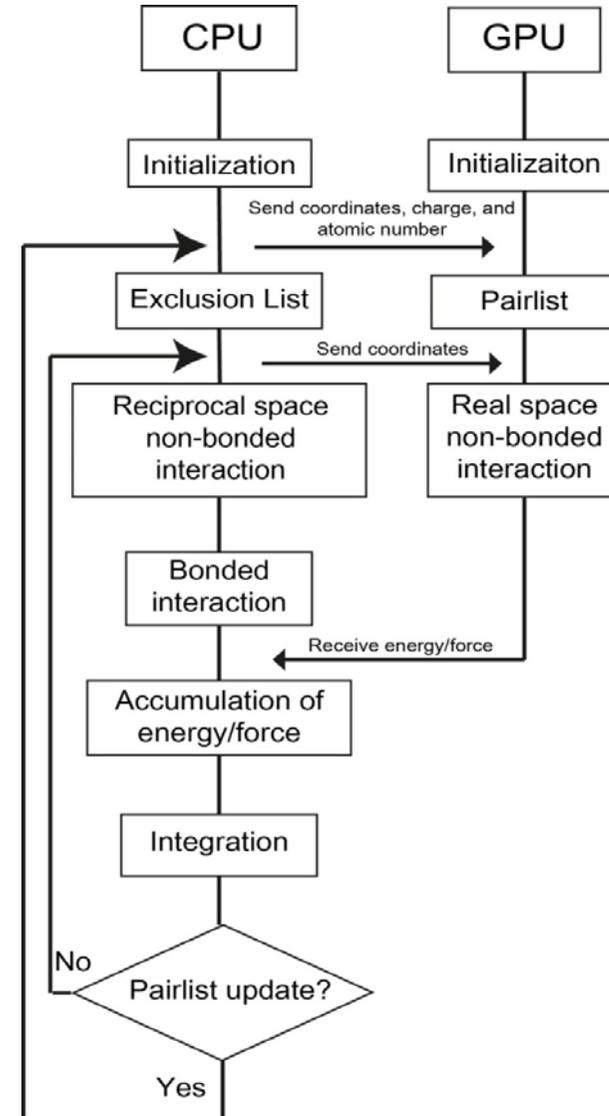
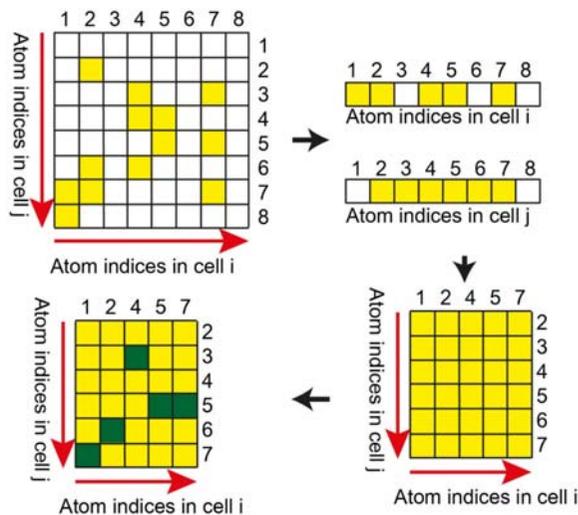


¹ KJ. Bowers et al, *JCP* 124, 184109 (2006)

GPGPUを用いた短距離相互作用計算の高速化

Jung et al., *J. Chem. Theory Comput.* 12:4947-4958 (2016)

- GPGPUを効果的に利用しつつ、並列効率の高い計算手法
 - ✓ 演算がメインとなる短距離相互作用計算はGPGPUを用いて高速に計算
 - ✓ 通信がメインとなる長距離相互作用計算はCPUを用いて計算
-
- GPUの小さいメモリ量に適したペアリストスキームの開発

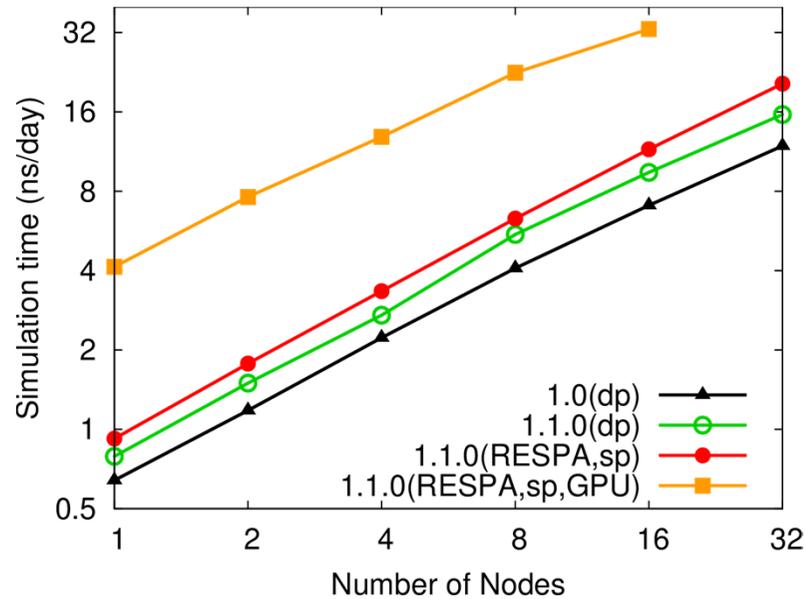


GENESIS 1.1の性能(ワークステーション)

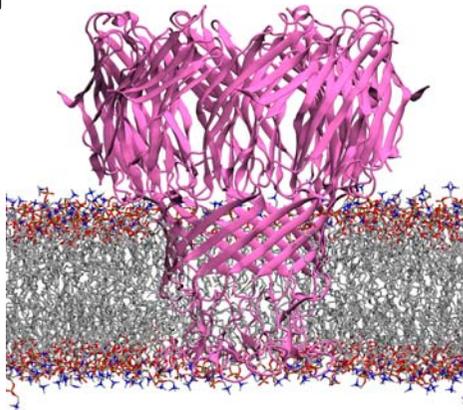
From http://www.aics.riken.jp/labs/cbrt/benchmark/#K_computer_benchmarks

Mycobacterium smegmatis porin A (MspA)

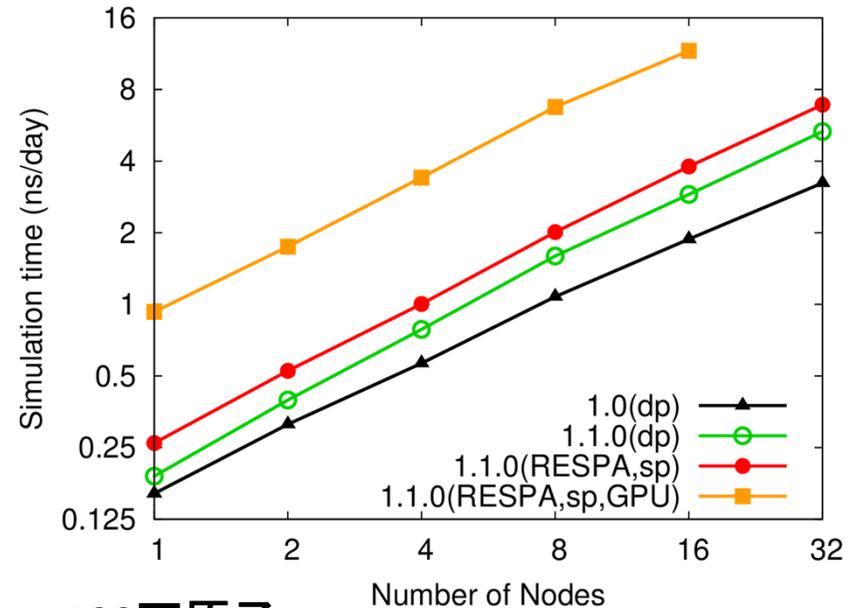
計算できる時間



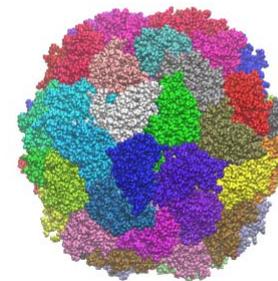
~21万原子



Satellite tobacco mosaic virus (STMV)



~100万原子

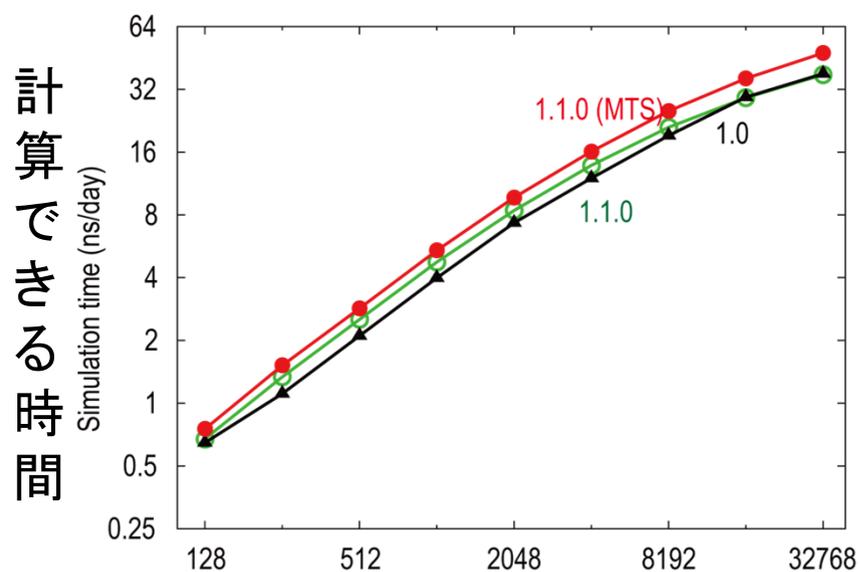


GENESIS 1.1の性能(京)

C. Kobayashi, J. Jung *et. al*, *J. Comput. Chem.* **38**, 2193-2206 (2017).

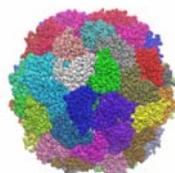


Satellite tobacco mosaic virus (STMV)

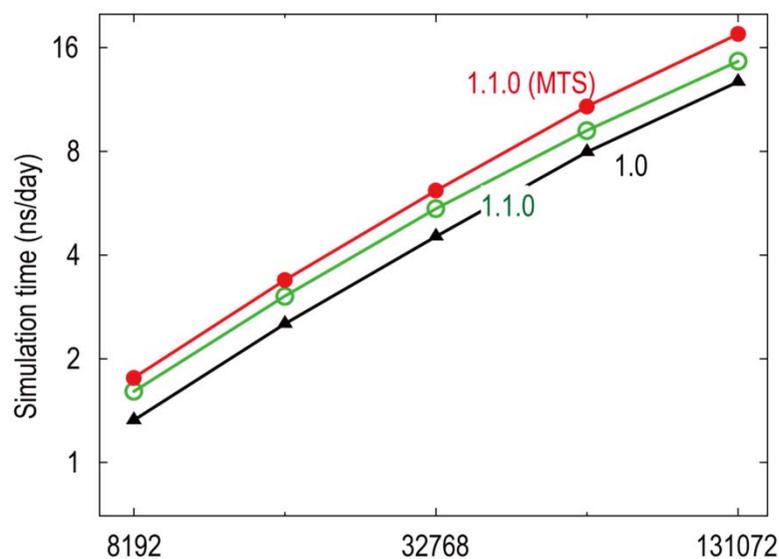


CPUコアの数

~100万原子

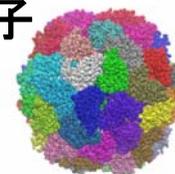


Satellite tobacco mosaic virus (STMV) × 27



CPUコアの数

~2800万原子



× 27

GENESISの内容物(1)

■ MDプログラム

- **ATDYN** (ATomic decomposition DYNamics simulator):
 - atomic decompositionを使用
 - 粗視化モデル(C α -GO, all atom GO)が計算可能
 - わかり易いコードでユーザによる開発可能
- **SPDYN** (SPatial decomposition DYNamics simulator):
 - domain decompositionを使用
 - 超並列・高速なルーチン (Midpoint cell/3次元分割FFT/GPU)

← 今回の実習ではこちらを使用

特徴	ATDYN	SPDYN
システムの分割法	原子分割	ドメイン分割
New lookup table	○	○
レプリカ交換法	○	○
粗視化モデル	○	×
3次元分割FFT	×	○
GPU	×	○
r-RESPA (1.1.0)	×	○

GENESISの内容物(2)

■ ATDYN/SPDYN共通

- 最適化
 - ◇ Steepest Decent法
- Integrator
 - ◇ Leapfrog
 - ◇ Velocity Verlet
- アンサンブル
 - ◇ NVE
 - ◇ NVT
 - Langevin
 - **Bussi**
 - Berendsen
 - ◇ NPT
 - Langevin Piston
 - **Bussi**
 - (Isotropy of Simulation box: Isotropic, Semi-iso, An-iso,

XY-fixed)

- 拘束計算(constraint)
 - ◇ SHAKE (Leapfrog)
 - ◇ RATTLE (Velocity Verlet)
 - ◇ SETTLE
- FFT (PME)
 - ◇ FFTE
- Restraint functions
 - ◇ Position
 - ◇ Bond
 - ◇ Angle
 - ◇ Dihedral angle
 - ◇ **RMSD**
- **Target/Steered MD**
- **String method**

(赤字は1.1.0の新機能)

GENESISの内容物(3)

■ その他解析ツール

- トラジェクトリ変換 (trjcnv)
 - ◇ crd_convert
 - ◇ pcrd_convert (並列I/O用計算用)
 - ◇ remd_convert (REMD計算用)
- トラジェクトリ解析 (trjana)
 - ◇ mbar_analysis
 - ◇ pathcv_analysis
 - ◇ pmf_analysis
 - ◇ gval_analysis
 - ◇ rmsd_analysis (RMSD、RMSF)
 - ◇ trj_analysis (距離、角度など)
 - ◇ wham_analysis
- リスタートファイル変換 (rstcnv)
 - ◇ rst_convert
 - ◇ rst_upgrade
- PCA解析 (pcaana)
 - ◇ avecrd_analysis
 - ◇ eigmat_analysis
 - ◇ flccrd_analysis
 - ◇ prjcrd_analysis
- 反応経路探索関連 (rpath_generator)
- 解析共通 (libana)

(赤字は1.1.0の新機能)