

計算生命科学の基礎VI

計算科学・データサイエンスと生命科学の融合：基礎から医療・創薬への応用まで

生体系の分子シミュレーションの新展開

2019年12月11日

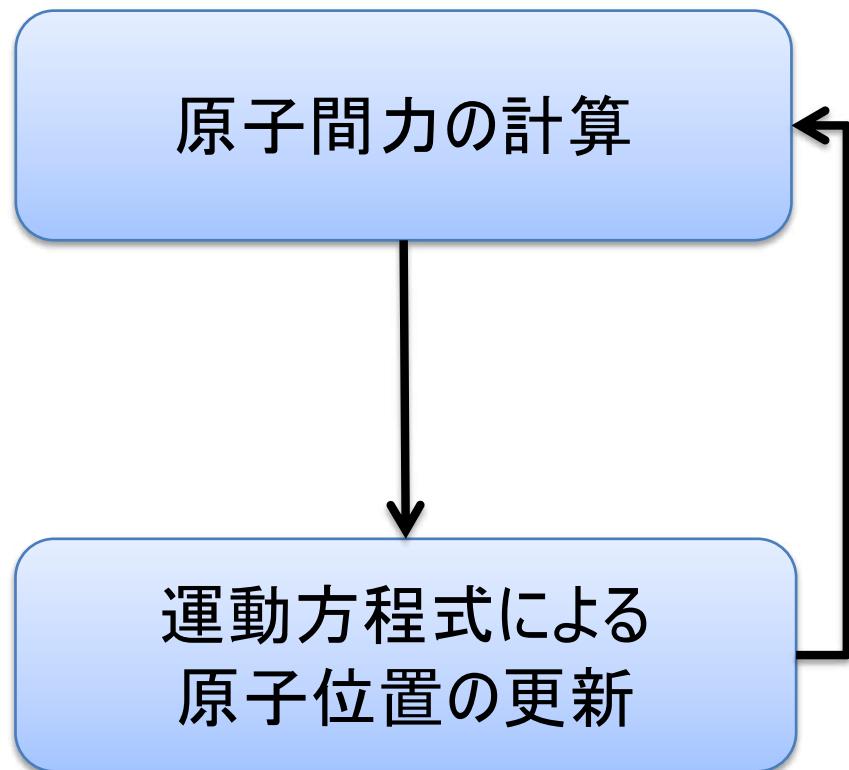
横浜市立大学 生命医科学研究科
理化学研究所 MIH

池口 満徳

今日の内容

- 分子動力学シミュレーションの新展開
 - 分子動力学シミュレーションとは
 - スパコン「京」から「富岳」へ
- 分子動力学シミュレーションの解析・応用
 - どのように分子動力学シミュレーションを活用するのか？
- 分子シミュレーションと人工知能(AI)の連携
 - LINCコンソーシアムの紹介
 - 最近の生体分子構造分野でのAIの活用

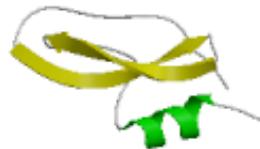
分子動力学シミュレーションの基本アルゴリズム



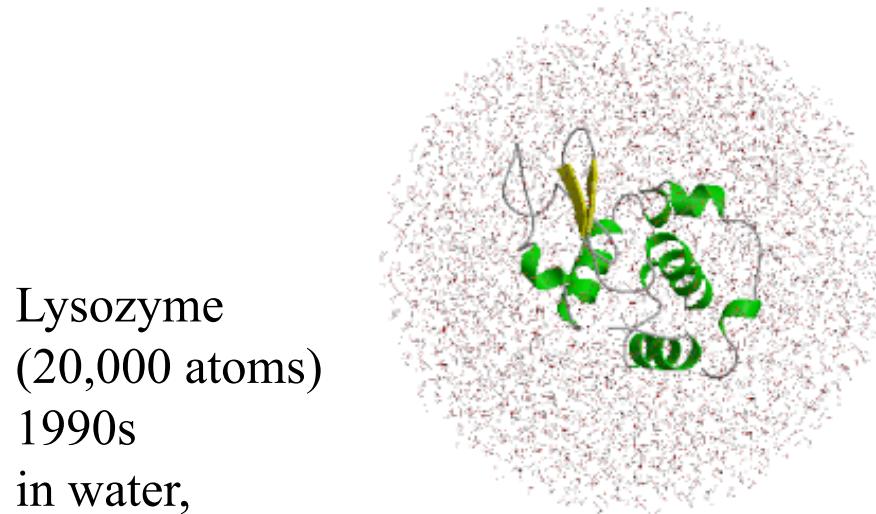
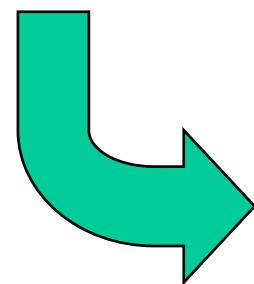
- ✓ 全原子MDでは、刻み幅は
1～2 fs (10^{-15} s) 程度
- ✓ 1ステップ1秒、計算にかかるとすると、
一日に、86,400ステップ=86 ps
しか計算できない！
- ✓ 1ステップ1ミリ秒で計算できると、
一日に、86,400,000ステップ=86 ns
- ✓ この速度でも1マイクロ秒のMD計算に
12日くらいかかる

生体分子の動きを扱うには、高速な計算が必須！

History of Molecular Dynamics Simulations of Proteins

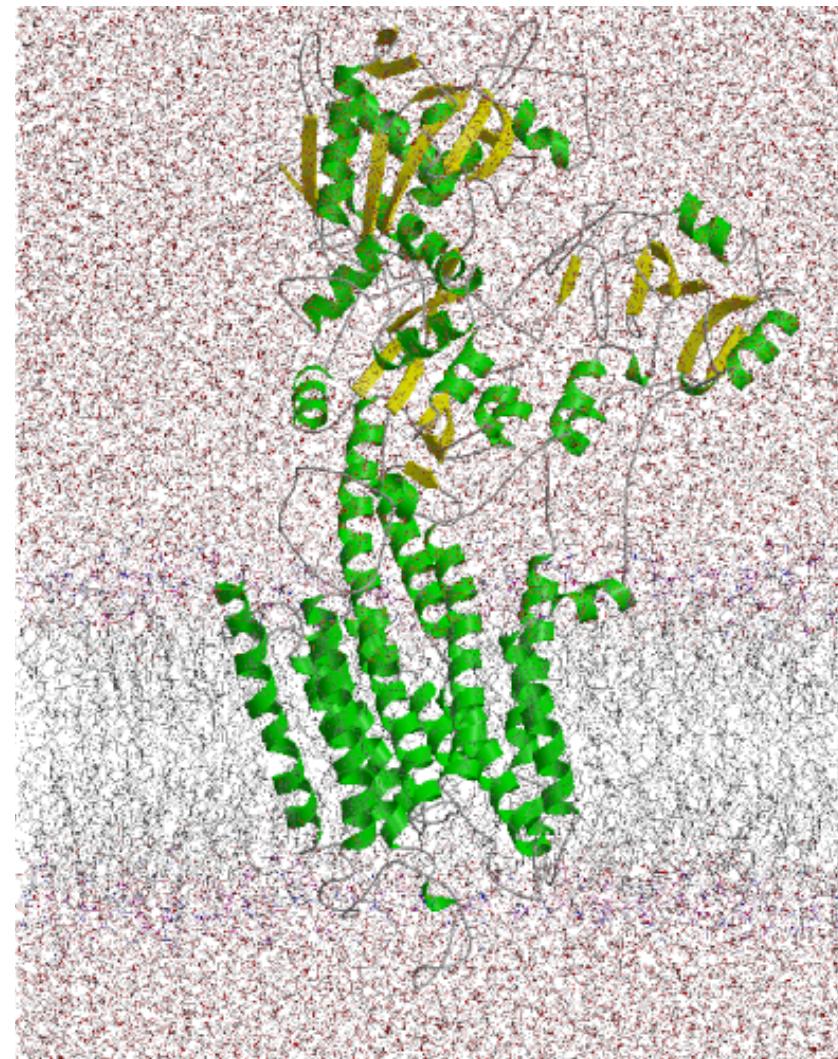


BPTI (3000 atoms)
Nature, 267, 585, 1977
in vacuum, 10 ps



Lysozyme
(20,000 atoms)
1990s
in water,
several ns

Ca²⁺-ATPase (more than 270,000 atoms including water)
2005 Sugita, Ikeguchi, Kidera, Toyoshima, JACS, 127, 6150

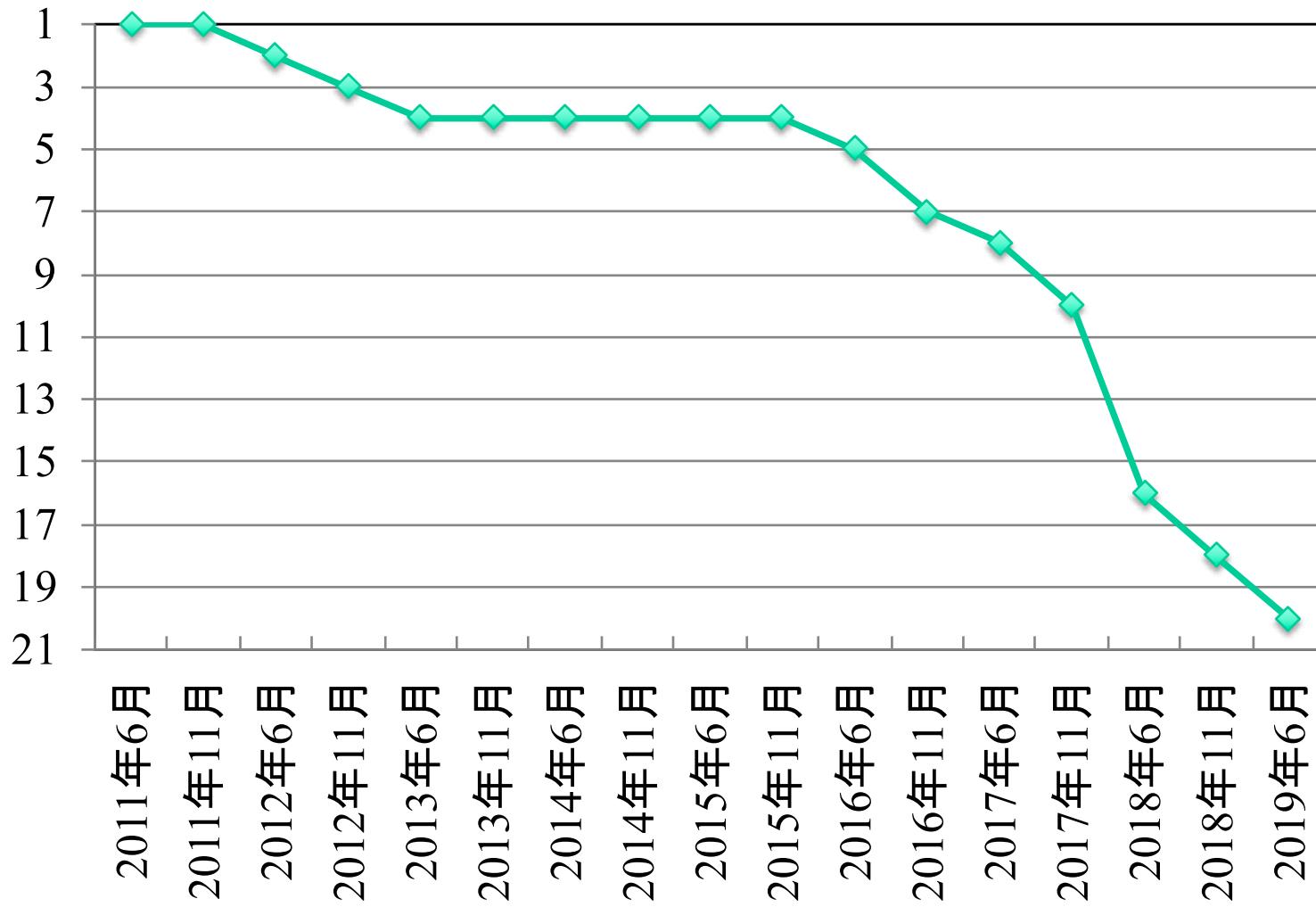


Flagship Supercomputer “K” 「京」



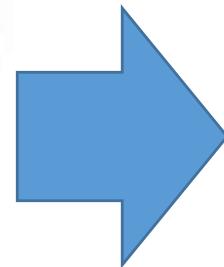
Ranking of K-Computer 「京」(Top500)

スノコソの世界順位





「京」 K-computer
663,552 cores
1.327 PB memory
11.28 PFLOPS (R_{peak})
10.51 PFLOPS (R_{max})



「富岳」

Fugaku

Shutdown in Aug. 2019

「富岳」のパワーを創薬に直結する研究体制

創薬は総合力が勝負。総合力で世界に勝つ！

ポスト京でのコ
デザイン

生体分子の
構造ダイナミクス

生体分子機能
の理解

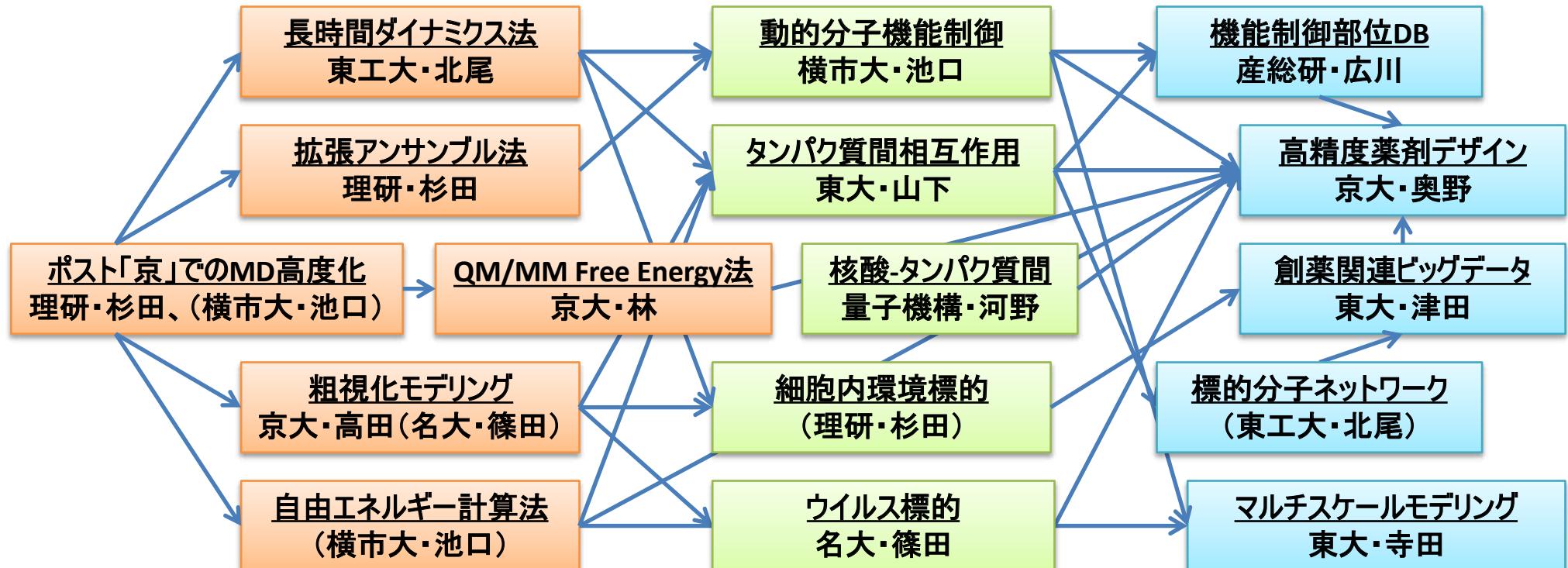
機能制御部位
の同定

候補化合物
デザイン

ポスト京を用いた超高速分
子シミュレーション

生体分子システムの時間的空間的機能解
析による次世代型創薬の開拓

ポスト京MDを機軸とした
革新的創薬計算基盤の構築



サブ課題A(理研・杉田):
ポスト京でのMD高度化と
アルゴリズム深化

サブ課題B(横市大・池口):
次世代創薬計算技術の開発

サブ課題C(京大・奥野):
創薬ビッグデータ統合システ
ムの開発

RIKEN Sugita Group: GENESIS

□ Performance Targets

- ✓ 100 times faster than K for some applications (tuning included)
- ✓ 30 to 40 MW power consumption

□ Peak Performance

	PostK	K
Peak DP (double precision)	400+ Pflops (34x +)	11.3 Pflops*
Peak SP (single precision)	800+ Pflops (70x +)	11.3 Pflops
Peak HP (half precision)	1600+ Pflops (141x +)	--
Total memory bandwidth	150+ PB/sec (29x +)	5,184TB/sec

* Reported in TOP500 (including I/O nodes)

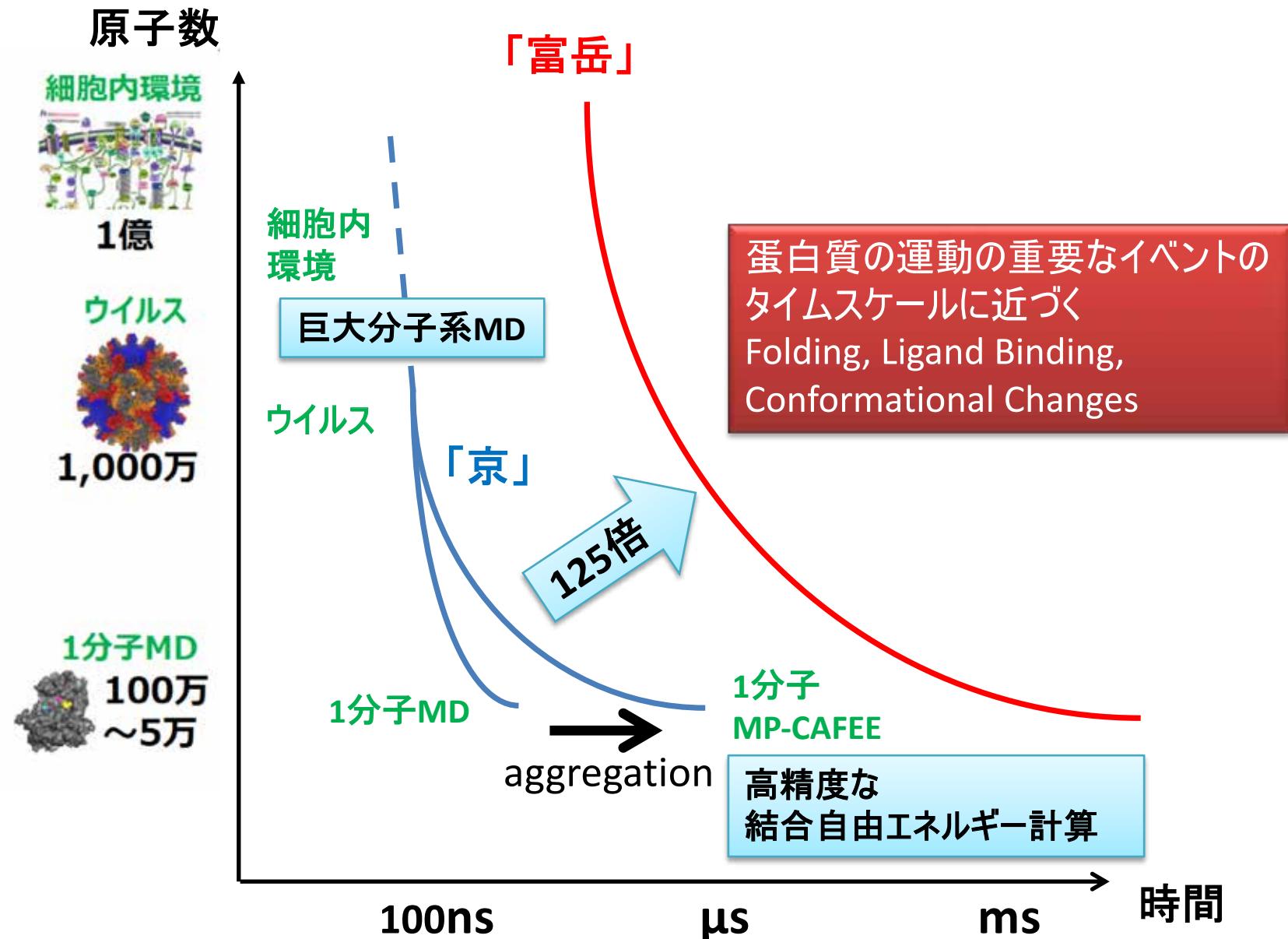
□ Geometric Mean of Performance Speedup of the 9 Target Applications over the K-Computer

37x +

□ Predicted Performance of 9 Target Applications As of 2019/05/14

Area	Priority Issue	Performance Speedup over K	Application	Brief description
Health and longevity Disaster prevention and Environment Energy issue Industrial competitiveness Basic science	1. Innovative computing infrastructure for drug discovery	125x +	GENESIS	MD for proteins
	2. Personalized and preventive medicine using big data	8x +	Genomon	Genome processing (Genome alignment)
	3. Integrated simulation systems induced by earthquake and tsunami	45x +	GAMERA	Earthquake simulator (FEM in unstructured & structured grid)
	4. Meteorological and global environmental prediction using big data	120x +	NICAM+ LETKF	Weather prediction system using Big data (structured grid stencil & ensemble Kalman filter)
	5. New technologies for energy creation, conversion / storage, and use	40x +	NTChem	Molecular electronic simulation (structure calculation)
	6. Accelerated development of innovative clean energy systems	35x +	Adventure	Computational Mechanics System for Large Scale Analysis and Design (unstructured grid)
	7. Creation of new functional devices and high-performance materials	30x +	RSDFT	Ab-initio simulation (density functional theory)
	8. Development of innovative design and production processes	25x +	FFB	Large Eddy Simulation (unstructured grid)
	9. Elucidation of the fundamental laws and evolution of the universe	25x +	LQCD	Lattice QCD simulation (structured grid Monte Carlo)

「京」から「富岳」へ：「富岳」で出来るようになること



汎用スパコンでの動的機能解析

数十 $\mu\text{s}/\text{day}$ 計算可能な専用計算機（ANTON2等）と違って、我々が使用可能なのは汎用スパコンである

一つの計算では、汎用計算機だと、現状からの飛躍的向上は、非常に困難である

しかし、汎用スパコンでは、ノード or CPUコアは沢山ある



レプリカを沢山用意して、統計的に計算を行うことで、実質的に、長時間ダイナミクス情報を得る

汎用スパコンでの動的機能解析

レプリカ系のMD計算

方法① 熱力学的方法

自由エネルギー分布を得る

(取りやすい複数の状態と頻度比を得る)

e.g. レプリカ交換、アンブレラサンプリング等

方法② 速度論的方法

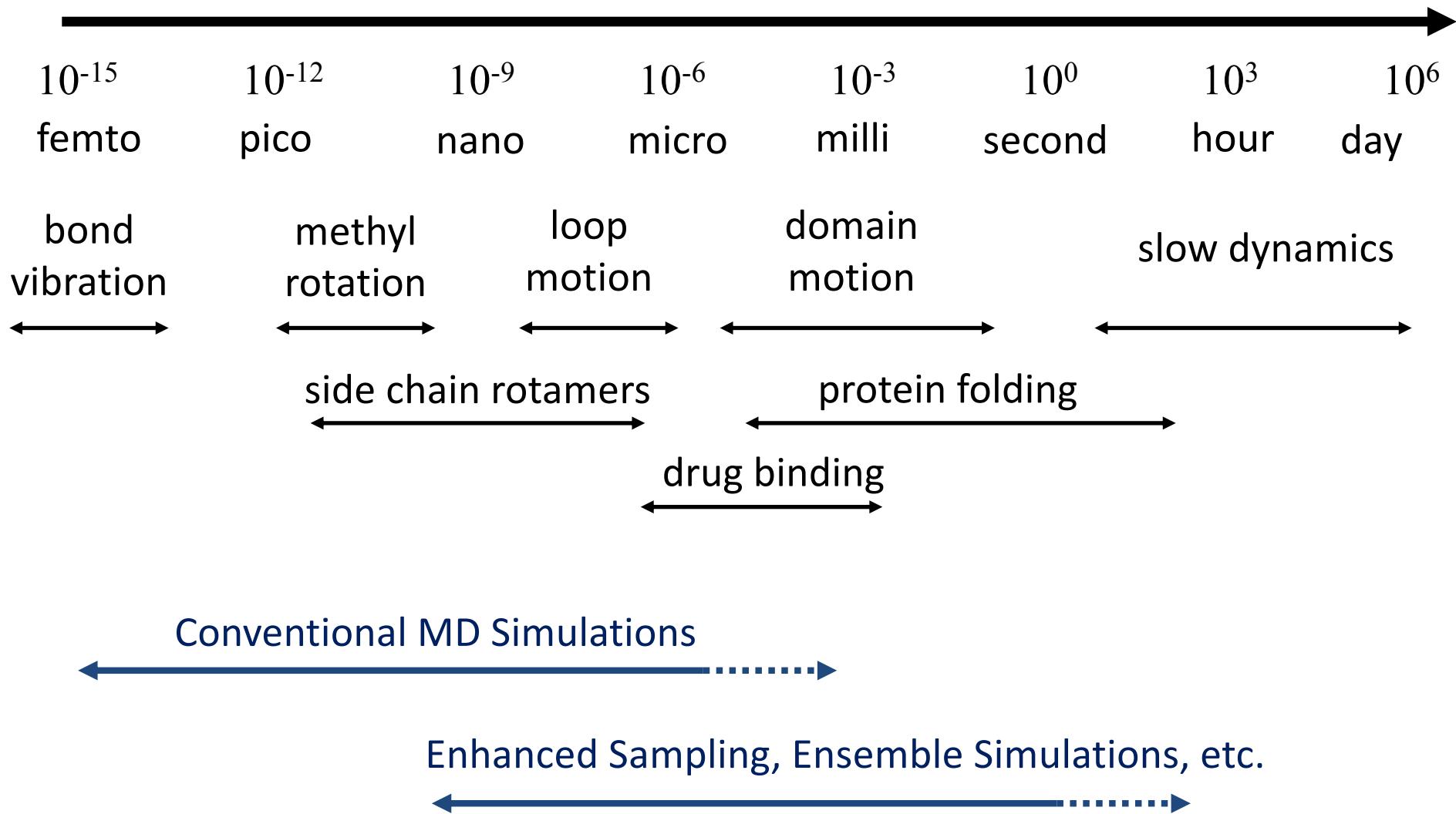
キネティックス（速度論）を得る

(取りやすい複数の状態間遷移を得る)

e.g. マルコフ状態遷移モデル (MSM)

Weighted Ensemble

Hierarchy of Protein Dynamics



MDで解析できること(例)

- 立体構造のゆらぎを観る
 - 蛋白質のどこが柔らかく、どこが堅いか？
- 運動の相関を観る
 - 蛋白質のどこが連動して運動しているか？
- 相互作用を観る
 - リガンドとの相互作用は、どこが強いか？
- 熱力学量を得る
 - 結合自由エネルギー、エンタルピー、エントロピーを得る
- リガンド結合ポーズを得る
 - リガンドの安定な結合ポーズ予測
- 立体構造変化を観る
 - どのように、蛋白質は立体構造変化するのか？
- 実験結果と一致する構造を得る
 - NMRやSAXSなどの実験結果と整合性を検討

MDで解析できること(例)

- 立体構造のゆらぎを観る
 - 蛋白質のどこが柔らかく、どこが堅いか？

MDシミュレーションの解析法: RMSD, RMSF

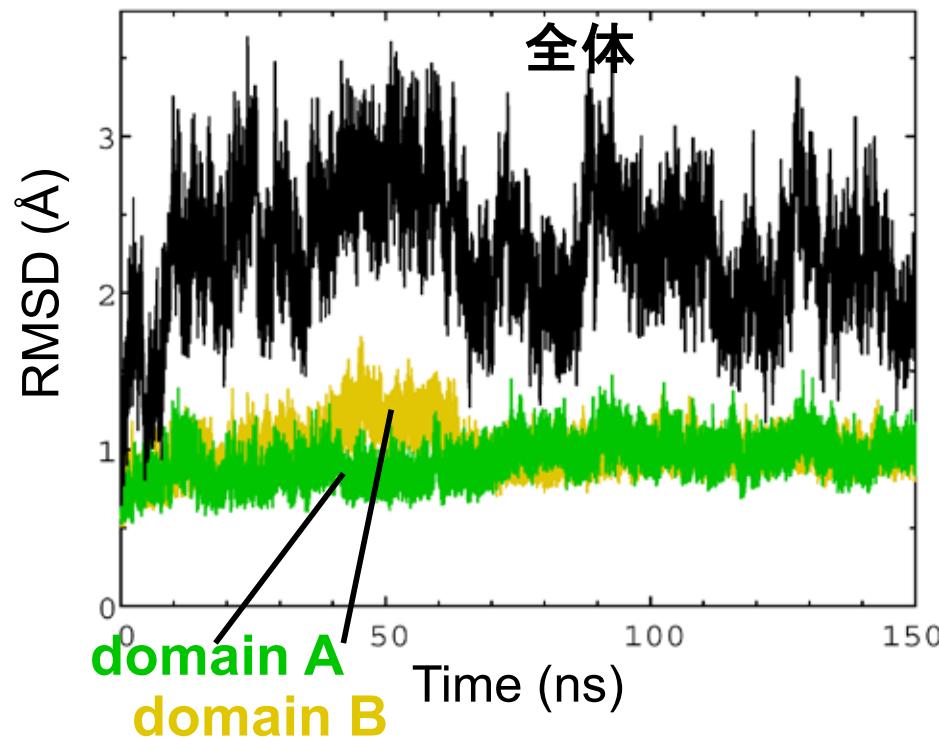
Root Mean Square Deviation (RMSD)

- ✓ 参照構造からのずれを観測
- ✓ 時間の関数
- ✓ MDの収束具合を見るのにも利用
- ✓ ドメイン運動している場合：
全体のRMSDは比較的大きくなる
各ドメインRMSDは小さい

$$\text{RMSD}(t) = \sqrt{\frac{(\sum_i x_i(t) - x_{\text{ref},i})^2}{N}}$$

MDシミュレーションの解析法 : RMSD

(RMSD from crystal structure)

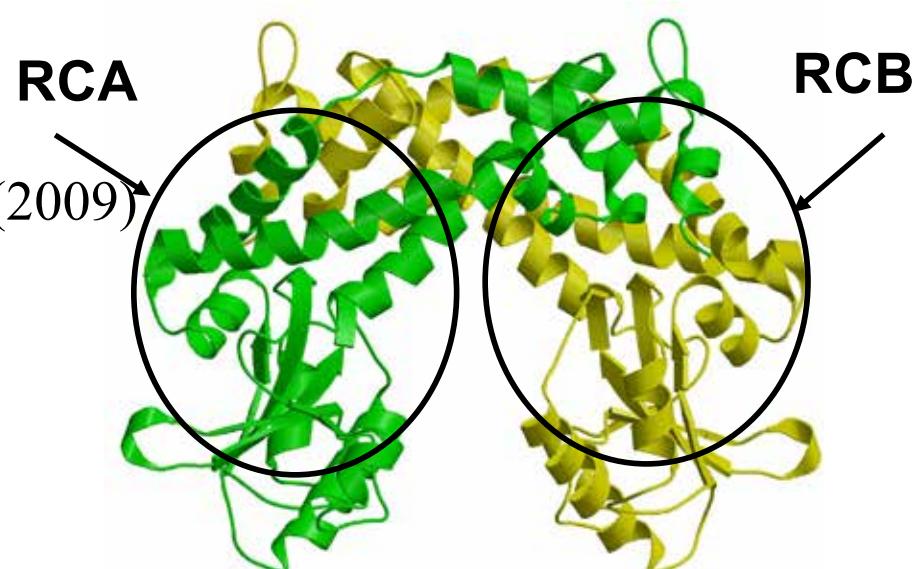


Root Mean Square Deviation (RMSD)

$$\text{RMSD} = \sqrt{\frac{(\sum_i \mathbf{x}_i - \mathbf{x}_{\text{ref } i})^2}{N}}$$

- ✓ 参照構造からのずれを観測
- ✓ MDの収束具合を見るのにも利用
- ✓ ドメイン運動している場合：
全体のRMSDは比較的大きくなる
各ドメインRMSDは小さい

Oroguchi, MI et al., Biophys. J., 96, 2808 (2009)



MDシミュレーションの解析法: RMSD, RMSF

Root Mean Square Deviation (RMSD)

- ✓ 参照構造からのずれを観測
- ✓ 時間の関数
- ✓ MDの収束具合を見るのにも利用
- ✓ ドメイン運動している場合:
全体のRMSDは比較的大きくなる
各ドメインRMSDは小さい

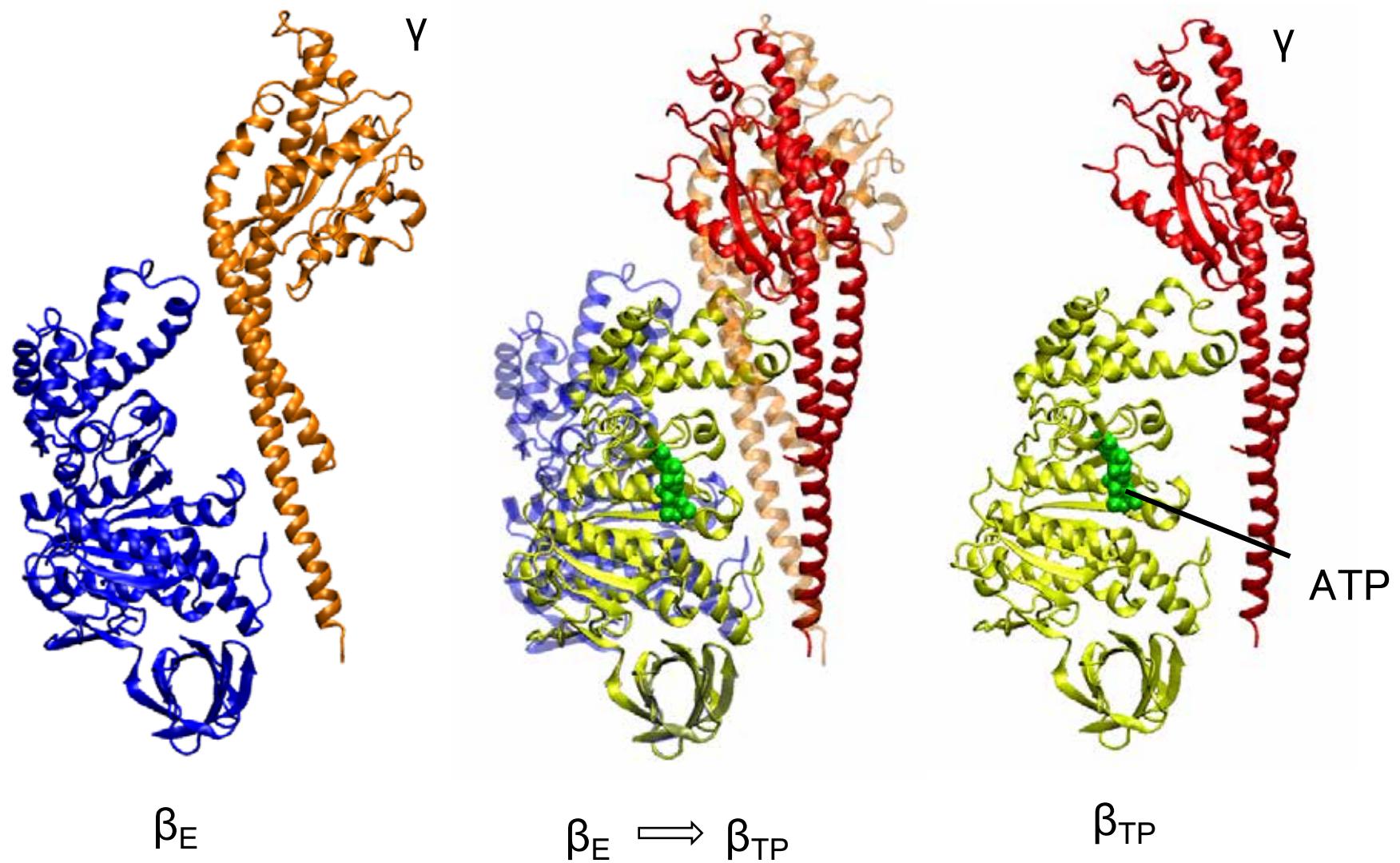
$$\text{RMSD}(t) = \sqrt{\frac{(\sum_i x_i(t) - x_{\text{ref},i})^2}{N}}$$

Root Mean Square Fluctuation (RMSF)

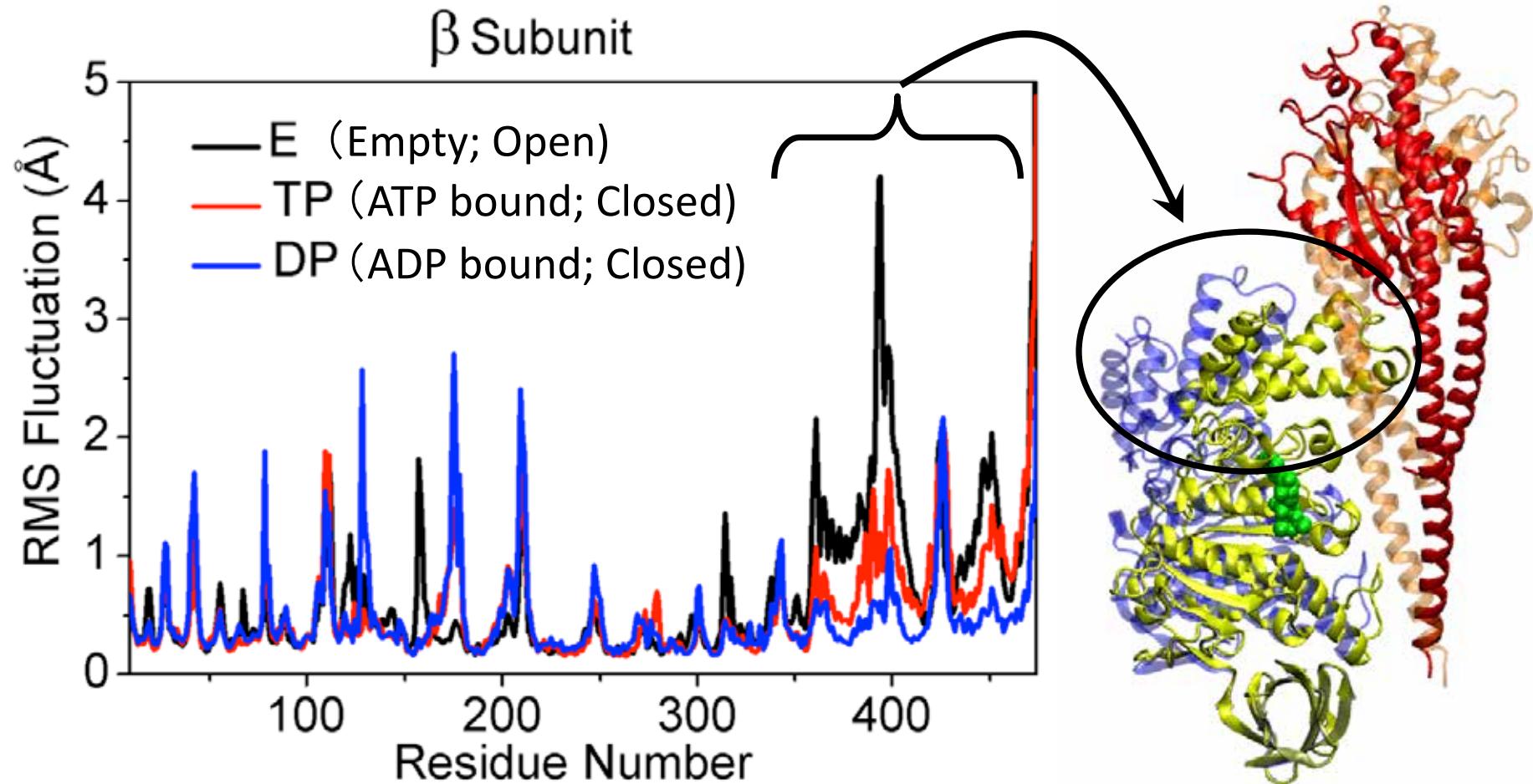
- ✓ 平均構造からのずれ(標準偏差)
- ✓ 原子番号・残基番号の関数
- ✓ 揺らいでいる箇所、固い箇所を判別
- ✓ 2次構造部位は揺らぎが小さい
- ✓ ループ部位は揺らぎが大きい

$$\text{RMSF}(i) = \sqrt{\frac{(\sum_t x_i(t) - x_{\text{ave},i})^2}{T}}$$

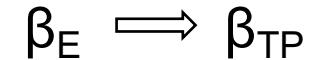
Conformational Changes of β subunit upon ATP binding



Protein Fluctuation (RMSF)



- ✓ Fluctuation of C-terminal domain: $\beta_E > \beta_{TP} > \beta_{DP}$
- ✓ Despite fairly similar conformations of β_{TP} and β_{DP} , fluctuations of C-terminal domain are different.



MDで解析できること(例)

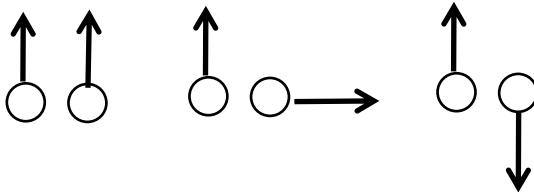
- 立体構造のゆらぎを観る
 - 蛋白質のどこが柔らかく、どこが堅いか？
- 運動の相関を観る
 - 蛋白質のどこが連動して運動しているか？

MDシミュレーションの解析法: 相関行列, PCA

Correlation Matrix (相関行列)

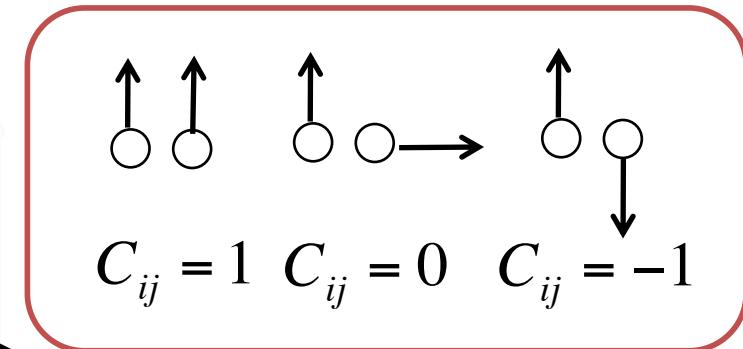
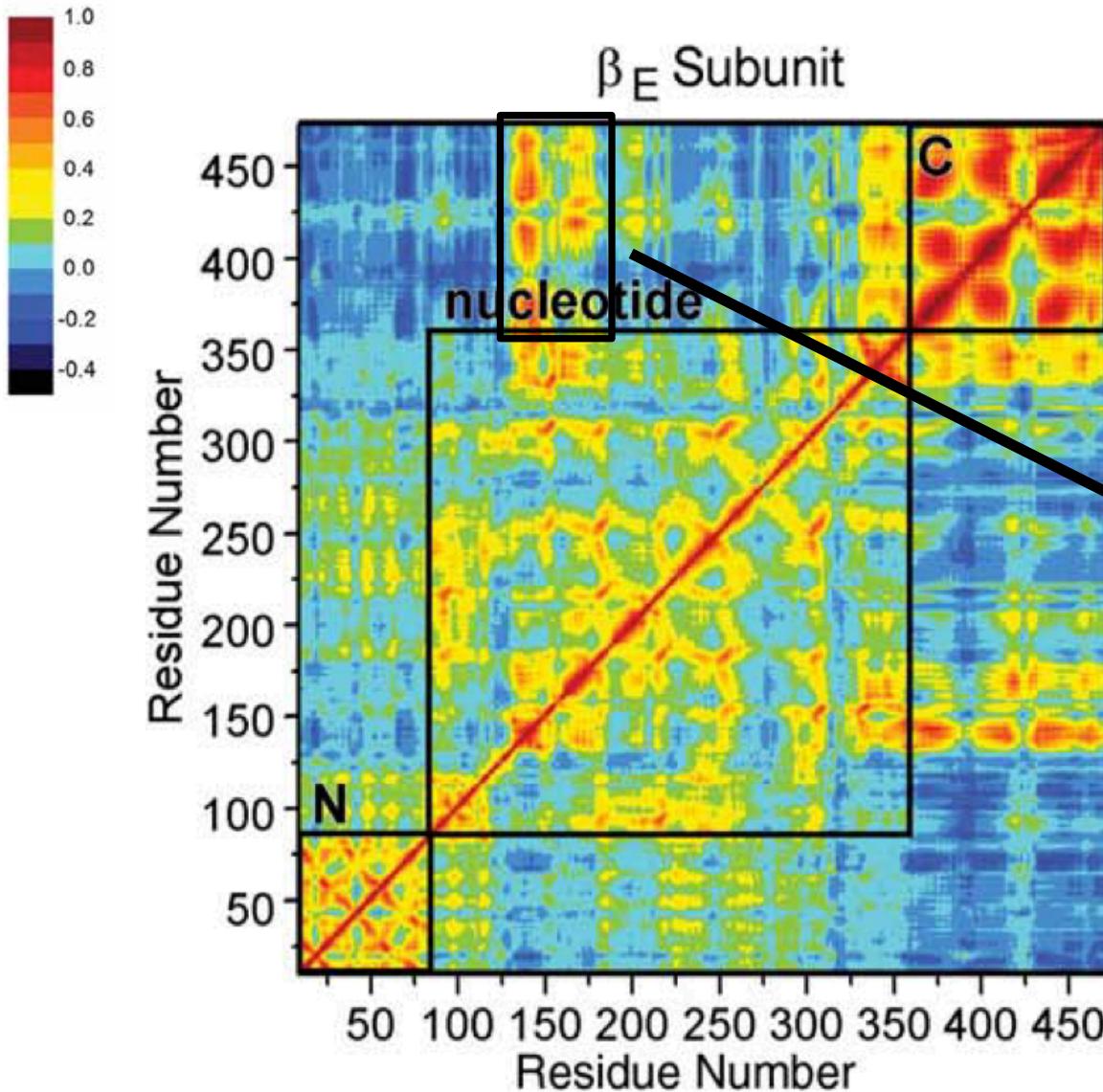
- ✓ 運動相関の度合
- ✓ 一緒に動いている箇所を判別
- ✓ 動的ドメインがわかる

$$C_{ij} = \frac{\langle \Delta\mathbf{r}_i \cdot \Delta\mathbf{r}_j \rangle}{\sqrt{\langle (\Delta\mathbf{r}_i)^2 \rangle \langle (\Delta\mathbf{r}_j)^2 \rangle}}$$



$$C_{ij} = 1 \quad C_{ij} = 0 \quad C_{ij} = -1$$

Correlated Motion in Protein



MDシミュレーションの解析法: 相関行列, PCA

Correlation Matrix (相関行列)

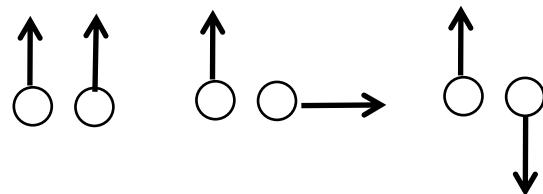
- ✓ 運動相関の度合
- ✓ 一緒に動いている箇所を判別
- ✓ 動的ドメインがわかる

Principal Component Analysis (PCA)

(主成分解析)

- ✓ 集団的運動を抽出
- ✓ 運動を主成分に分解
- ✓ 履歴を主成分方向へ射影することで運動の大まかな遷移がわかる

$$C_{ij} = \frac{\langle \Delta\mathbf{r}_i \cdot \Delta\mathbf{r}_j \rangle}{\sqrt{\langle \Delta\mathbf{r}_i^2 \rangle \langle \Delta\mathbf{r}_j^2 \rangle}}$$



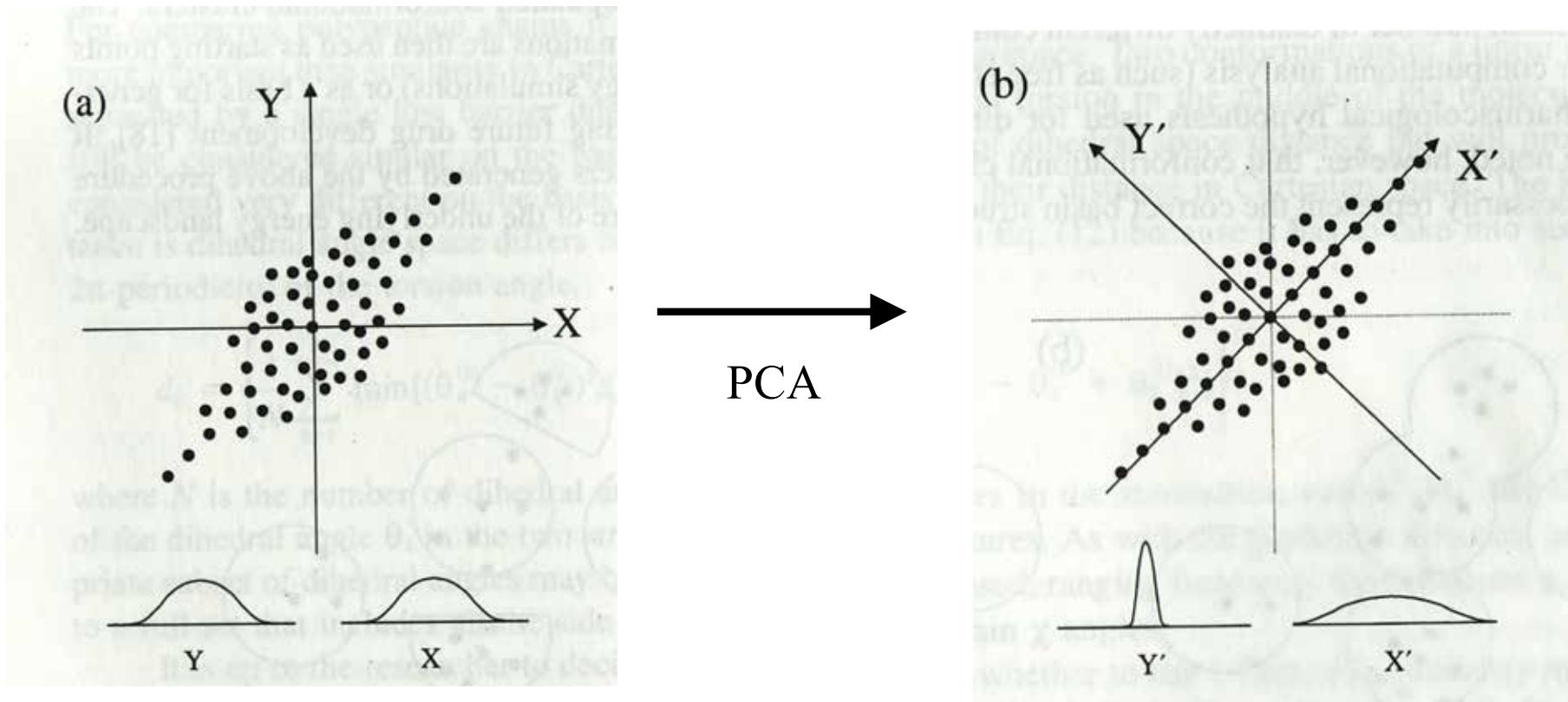
$$C_{ij} = 1 \quad C_{ij} = 0 \quad C_{ij} = -1$$

分散共分散行列を対角化

$$\langle \Delta x_i \Delta x_j \rangle$$
$$i = 1 \dots 3N$$

MDシミュレーションの解析法: PCA

主成分解析 Principal Component Analysis (PCA)

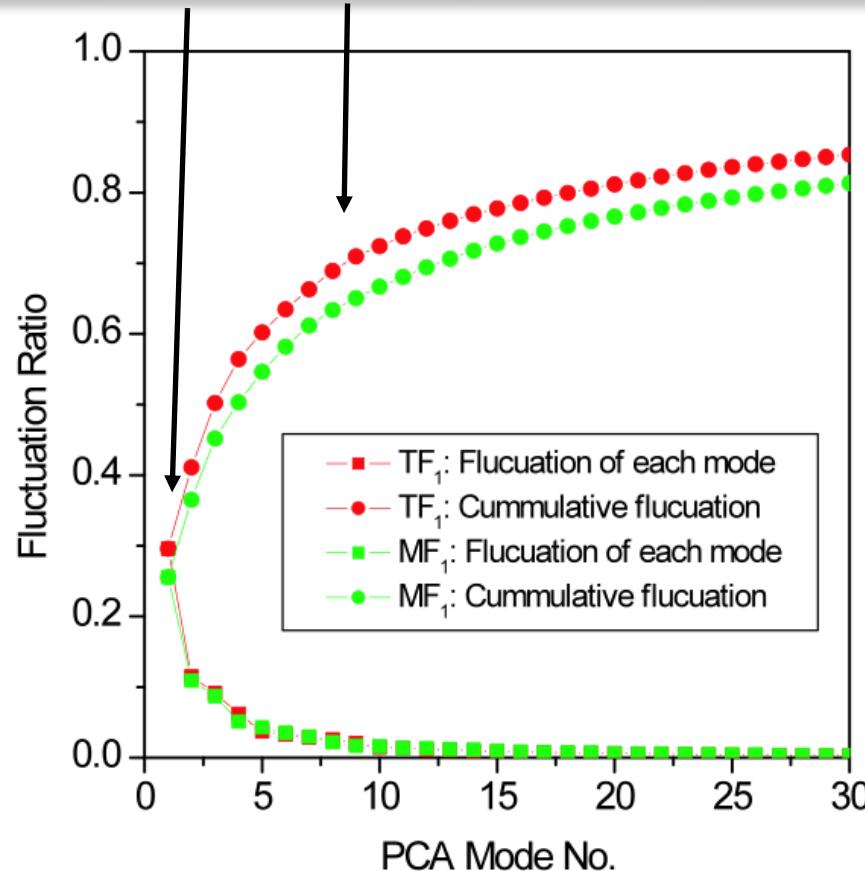


- ✓ PCAでは、大きな揺らぎ方向を抽出
- ✓ 原子の集団運動を得ることができる。

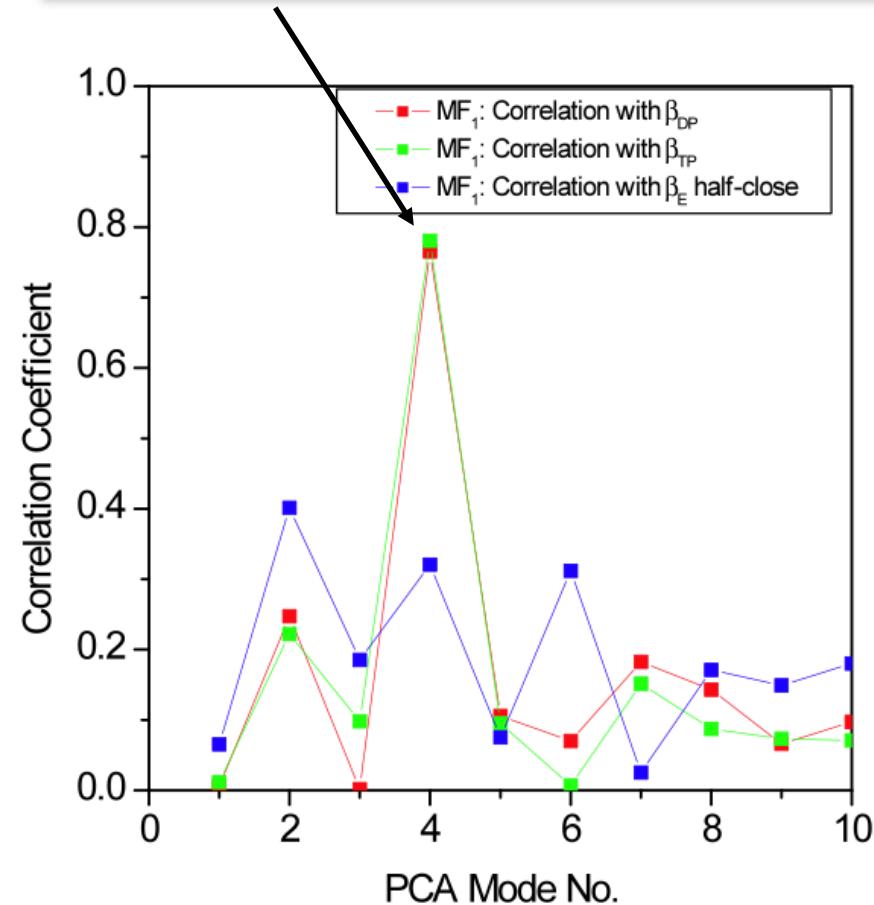
Computational Biochemistry and Biophysics, Becker et al. eds., 2001.

PCA: F₁-ATPase βサブユニット(単体)

- ✓ Top mode は、全体揺らぎの 30 % を占める
- ✓ Top ten modes は、全体揺らぎの 70 % を占める



- ✓ 4番目のモードは ATP結合時の構造変化と良い相関を示す

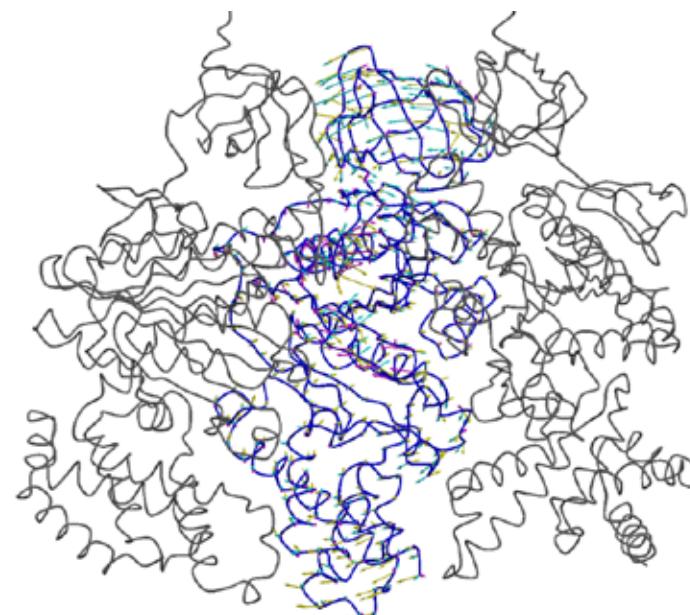
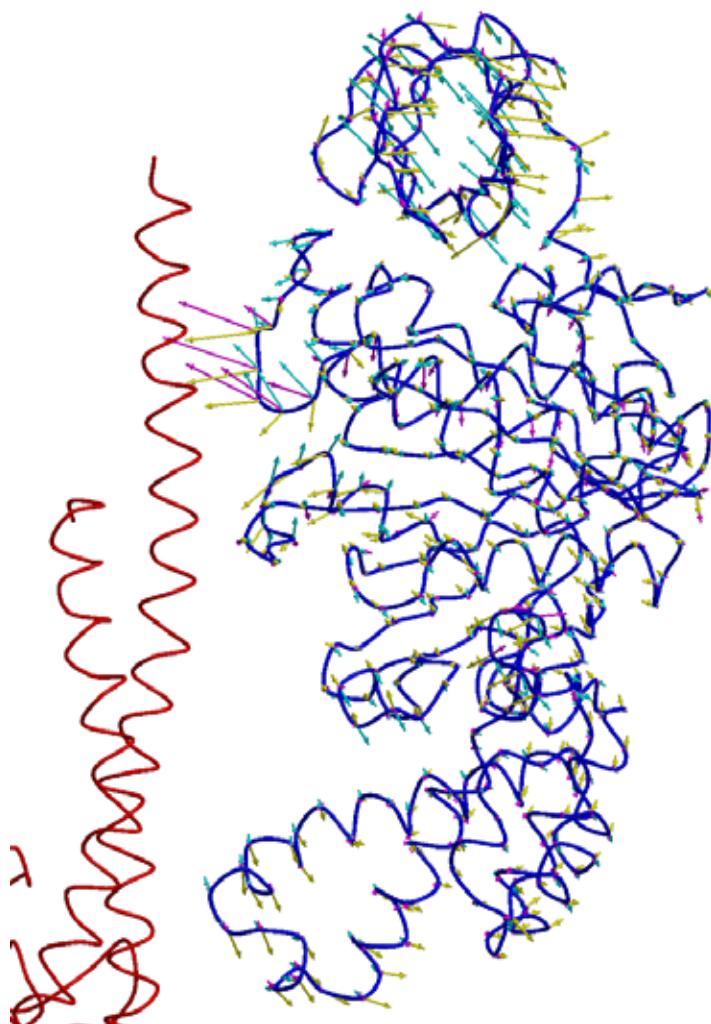


蛋白質は、非常に大きな自由度を持つが、揺らぎでみると、異方性が強く、限られた方向にのみ、揺らいでいる傾向がある。

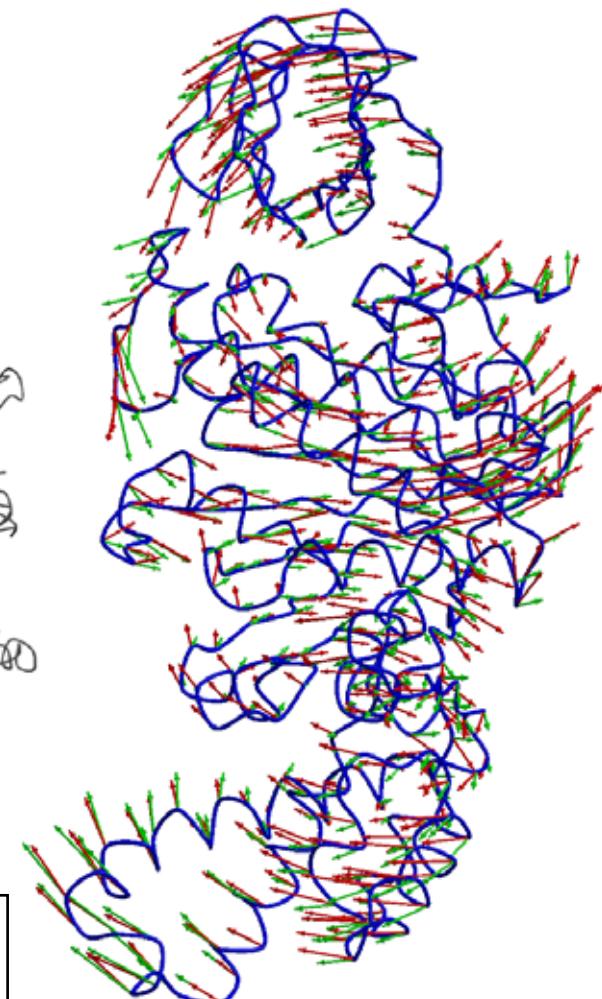
Principal Component Analysis: Mode 1-4

- ✓ Relaxation from interactions with adjacent subunits

Mode 1: magenta
Mode 2: yellow
Mode 3: cyan



Mode 4: green
 $\beta_E \rightarrow \beta_{DP}$: red



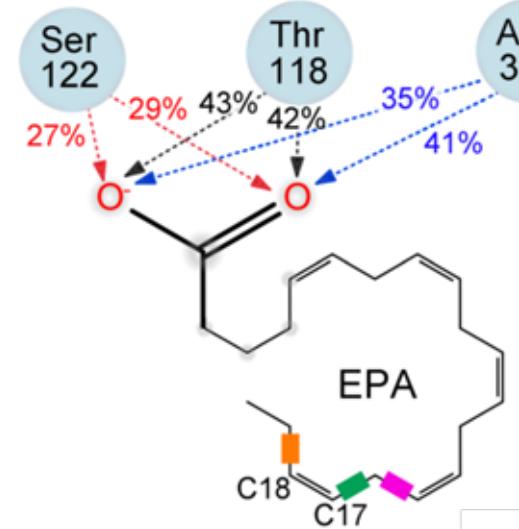
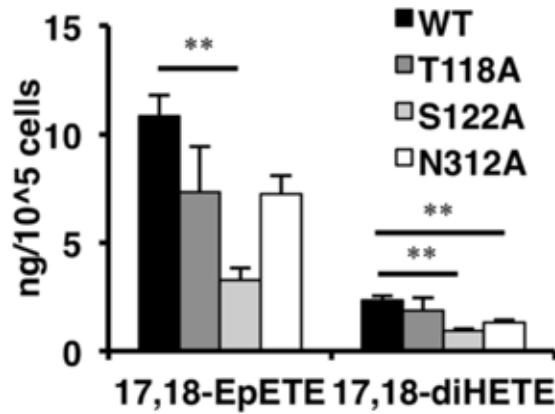
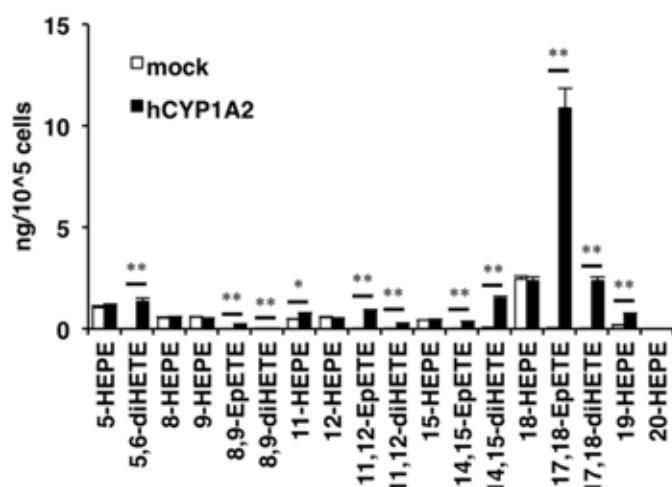
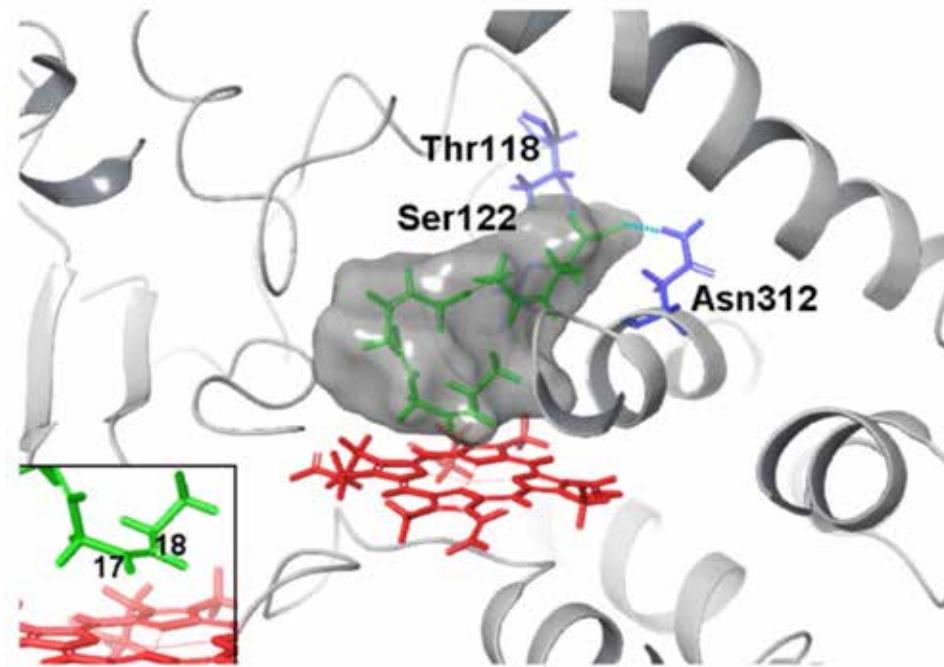
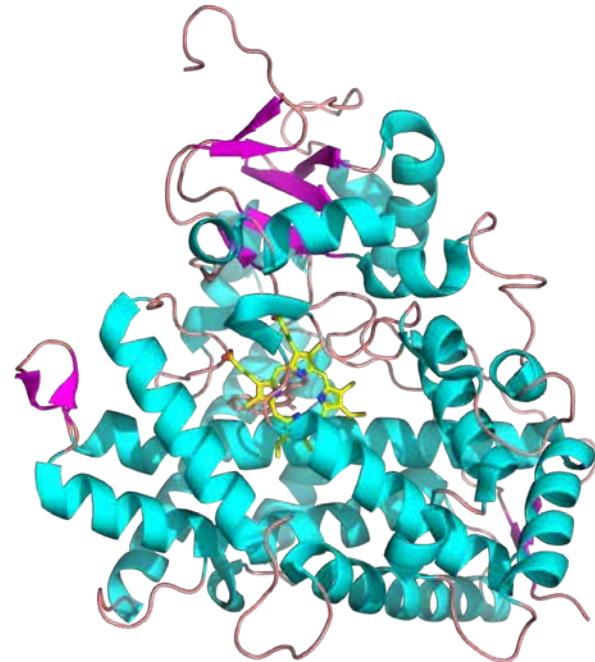
- ✓ Intrinsic flexibility is well correlated with structural transition

MDで解析できること(例)

- 立体構造のゆらぎを観る
 - 蛋白質のどこが柔らかく、どこが堅いか？
- 運動の相関を観る
 - 蛋白質のどこが連動して運動しているか？
- 相互作用を観る
 - リガンドとの相互作用は、どこが強いか？

Interaction between ligand and protein

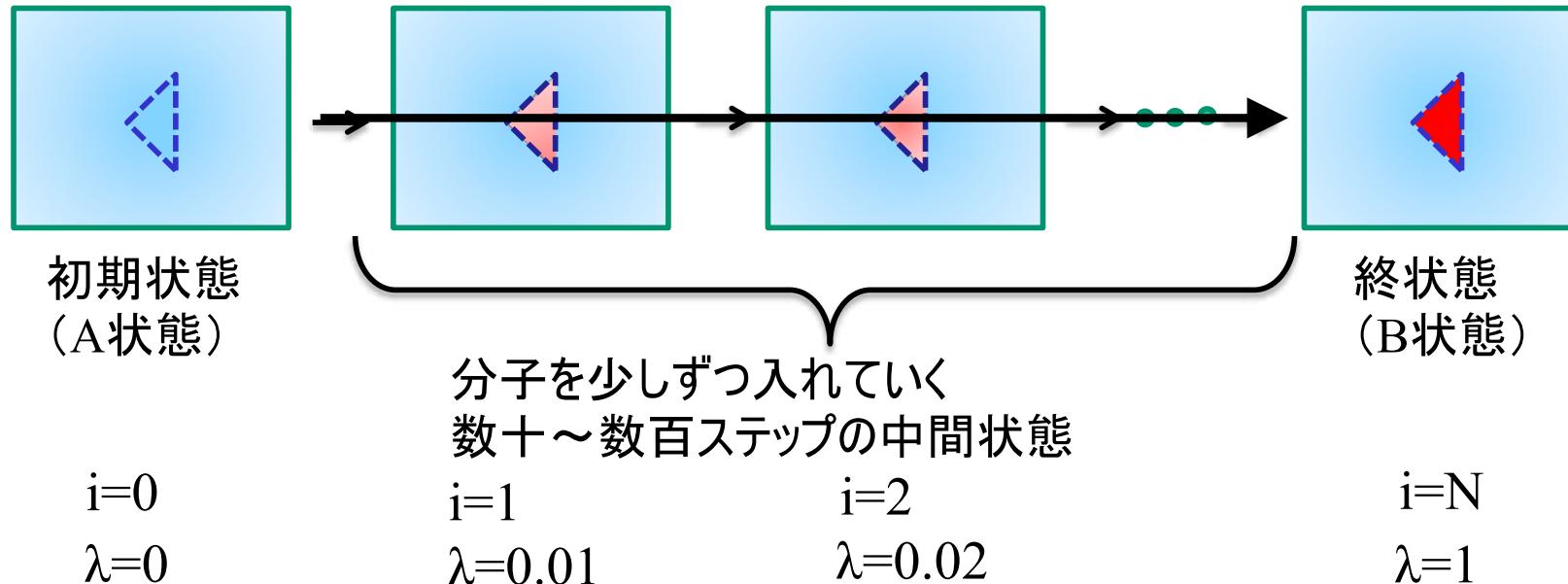
Interaction between CYP1A2 and EPA



MDで解析できること(例)

- 立体構造のゆらぎを観る
 - 蛋白質のどこが柔らかく、どこが堅いか？
- 運動の相関を観る
 - 蛋白質のどこが運動して運動しているか？
- 相互作用を観る
 - リガンドとの相互作用は、どこが強いか？
- 熱力学量を得る
 - 結合自由エネルギー、エンタルピー、エントロピーを得る

MDによる自由エネルギー評価法



初期(A)状態から終(B)状態に変わるときの自由エネルギー差

$$\Delta G = -k_B T \sum_i^{N-1} \ln \left\langle \exp \left(-\frac{U_{i+1} - U_i}{k_B T} \right) \right\rangle_i$$

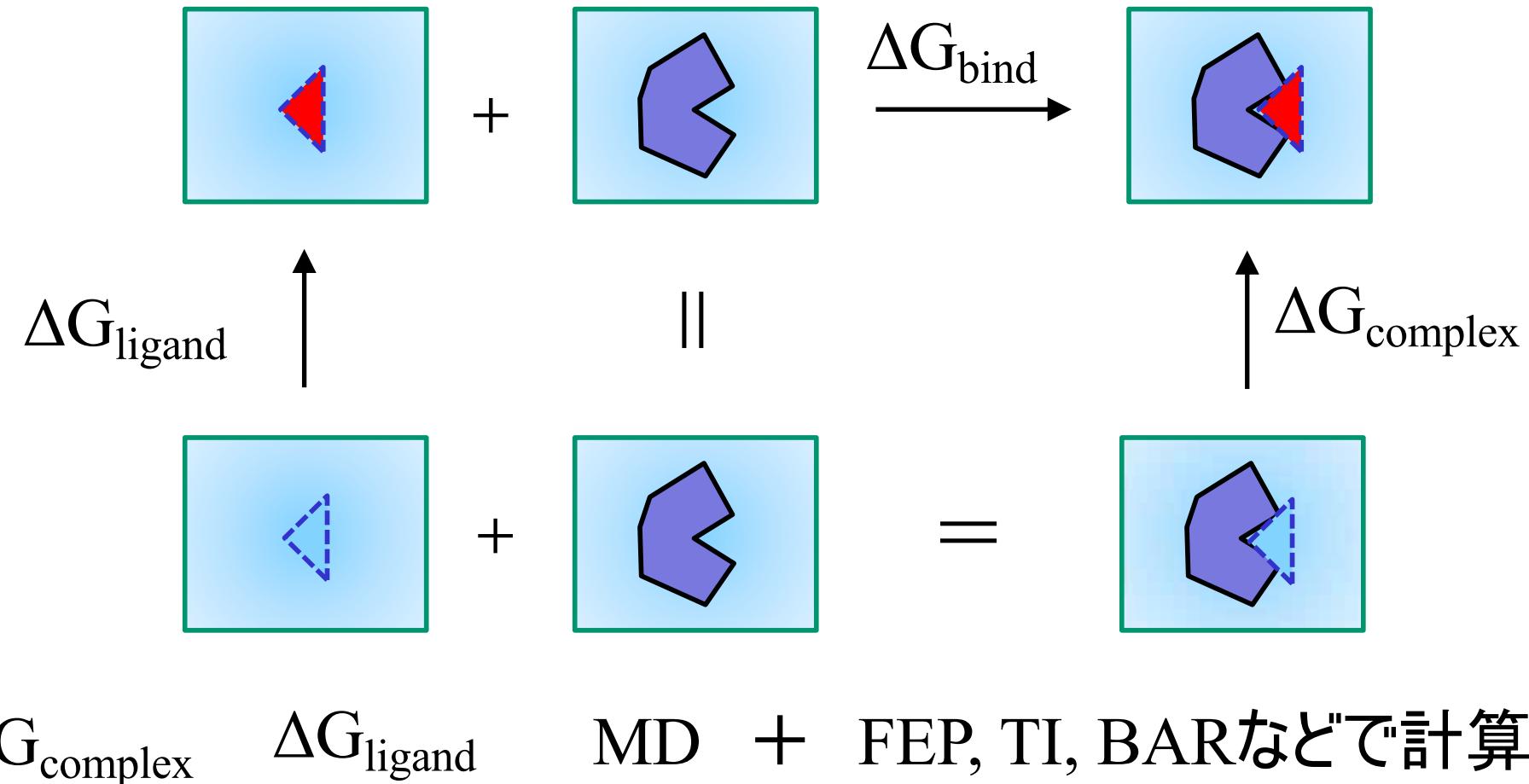
自由エネルギー擾動法
Free Energy Perturbation

$$= \int_0^1 \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda d\lambda$$

熱力学積分法 Thermodynamic Integration

この他にも Bennett Acceptance Ratio (BAR)など

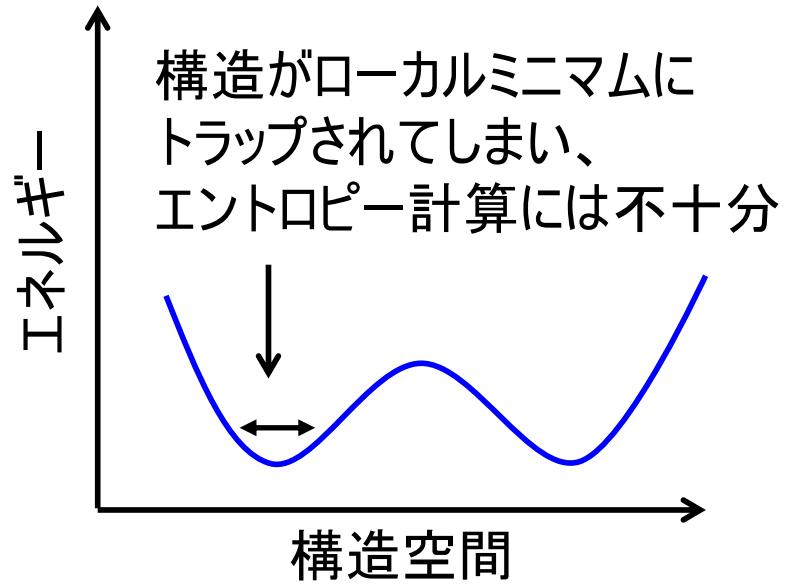
熱力学サイクル



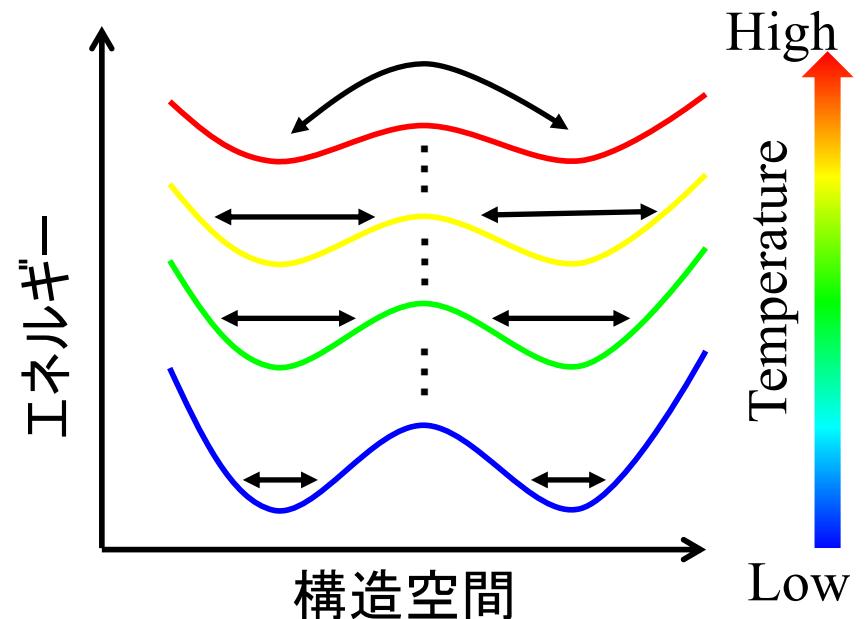
$$\Delta G_{\text{bind}} = \Delta G_{\text{complex}} - \Delta G_{\text{ligand}}$$

温度レプリカ交換MD法

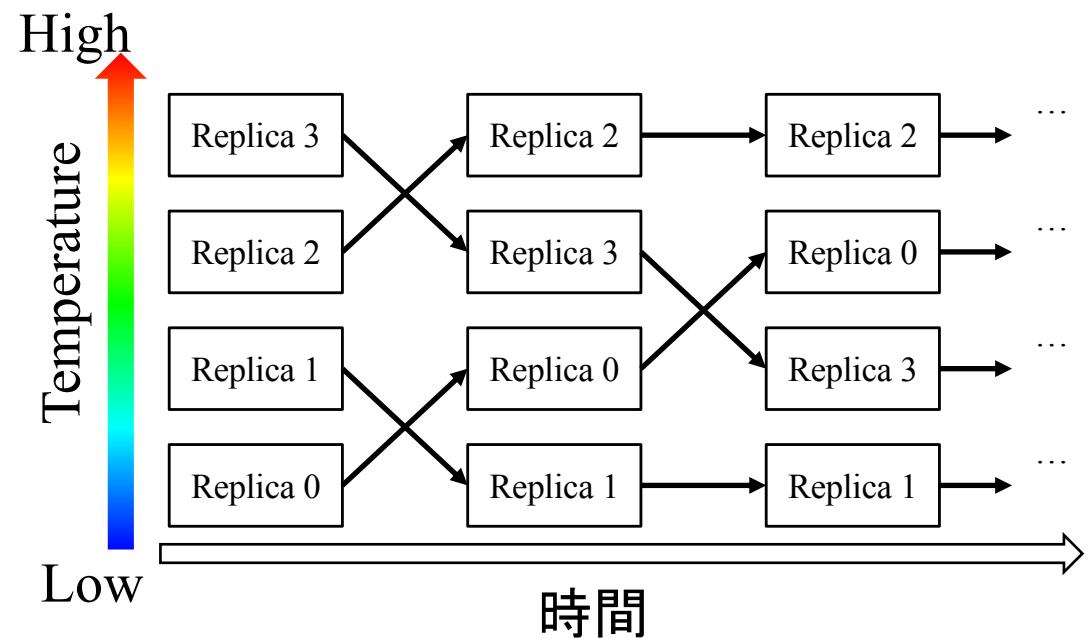
通常のMD法



温度レプリカ交換MD法

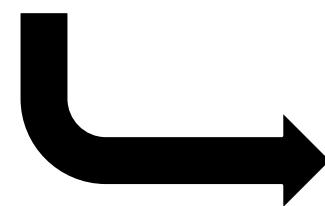
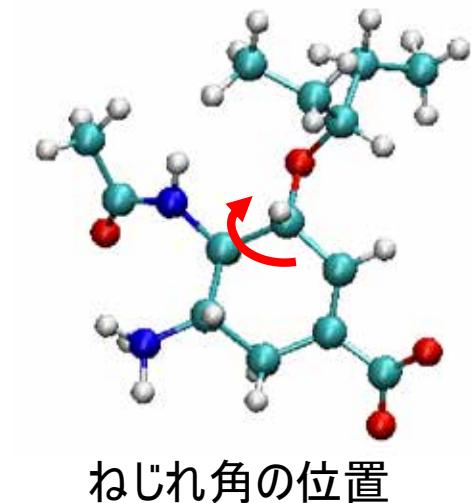
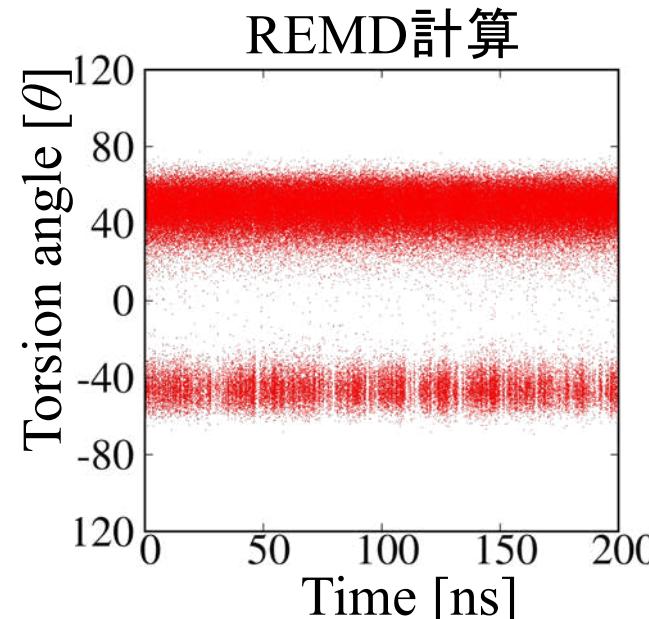
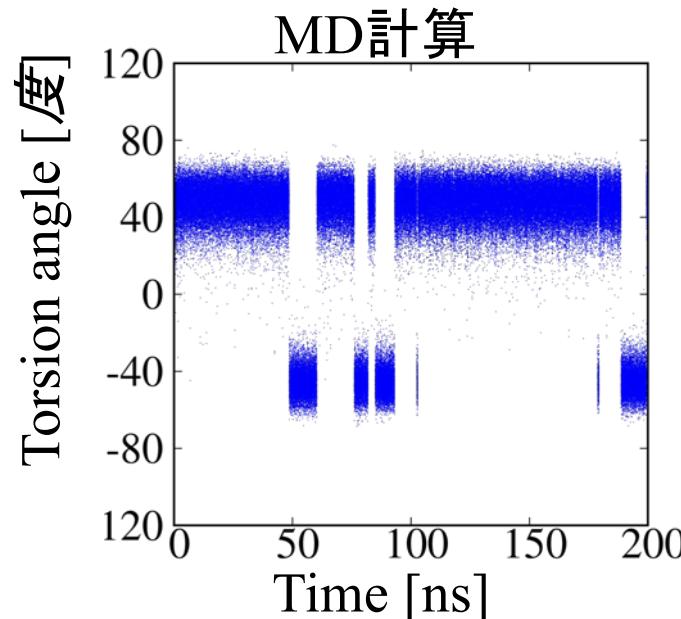


温度レプリカ交換MD法の導入により、幅広い構造サンプリングが可能になる

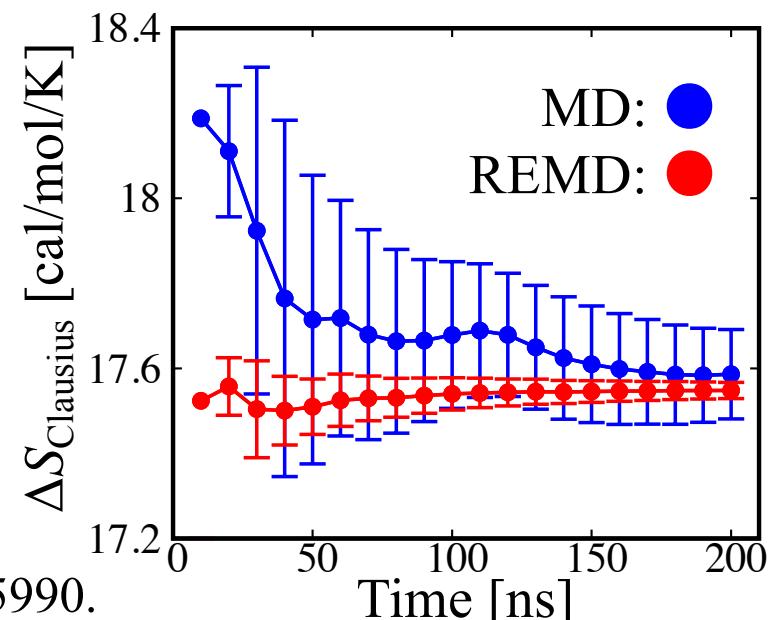


温度レプリカ交換MD(REMD)法により計算の収束性が向上

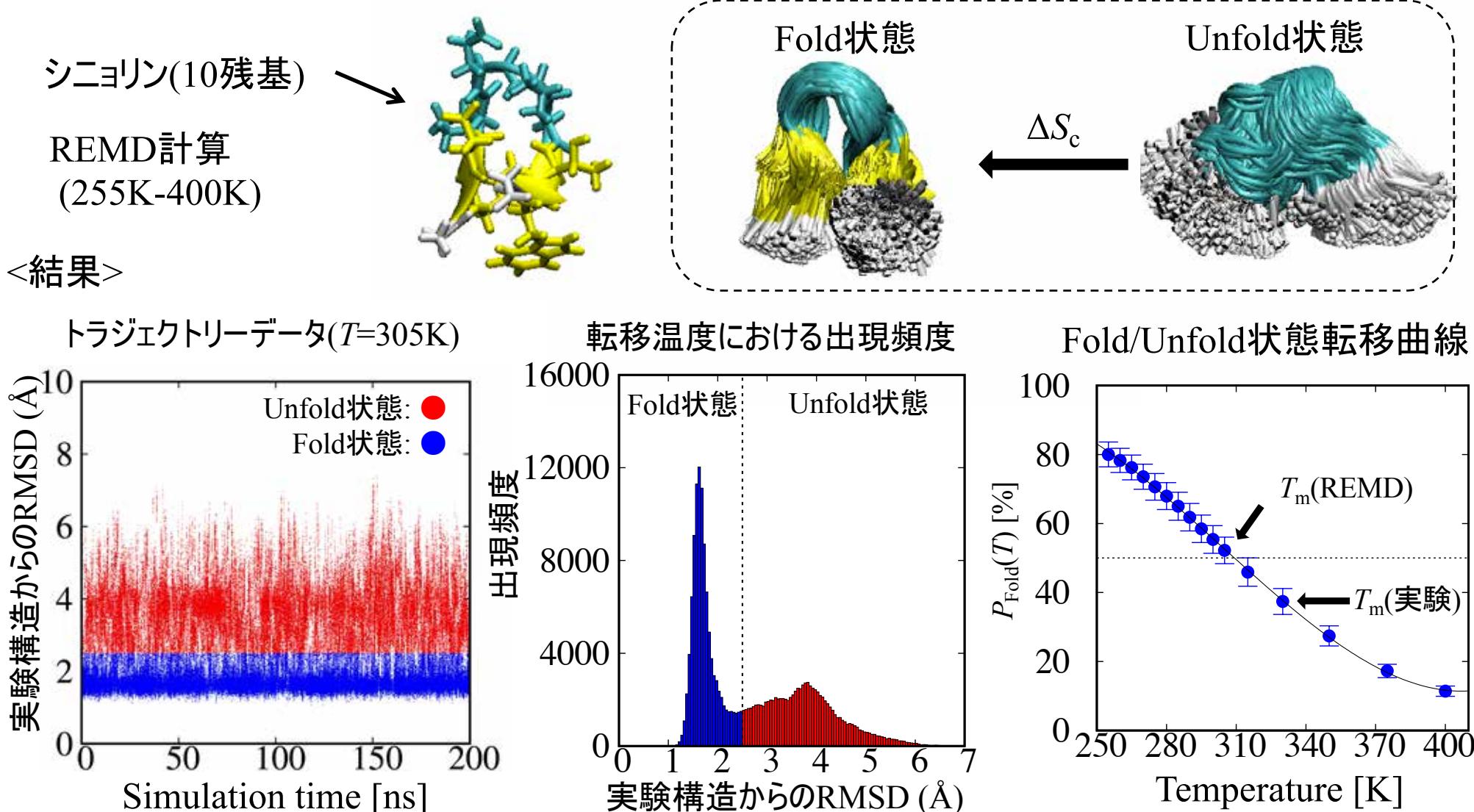
例: Oseltamivir シミュレーション中のねじれ角の分布



温度レプリカ交換MDの導入により、
構造分布の収束性が向上し、
構造エントロピー計算の収束性も向上

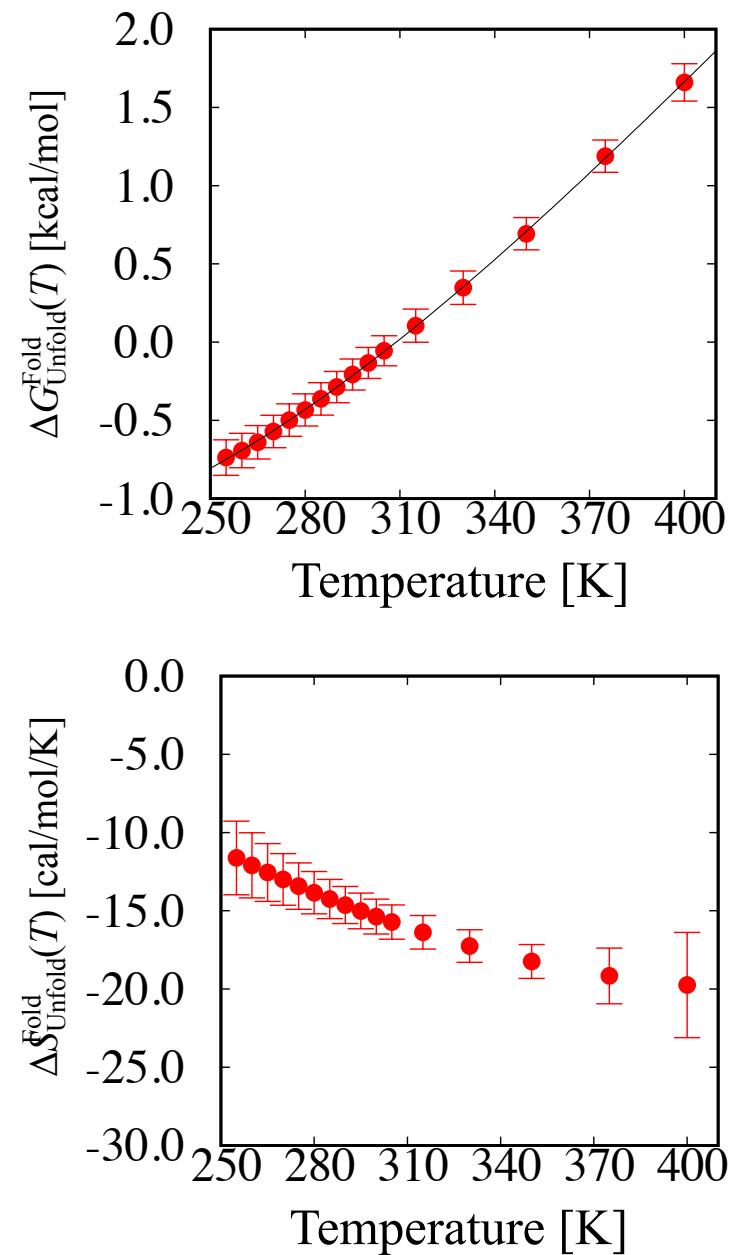
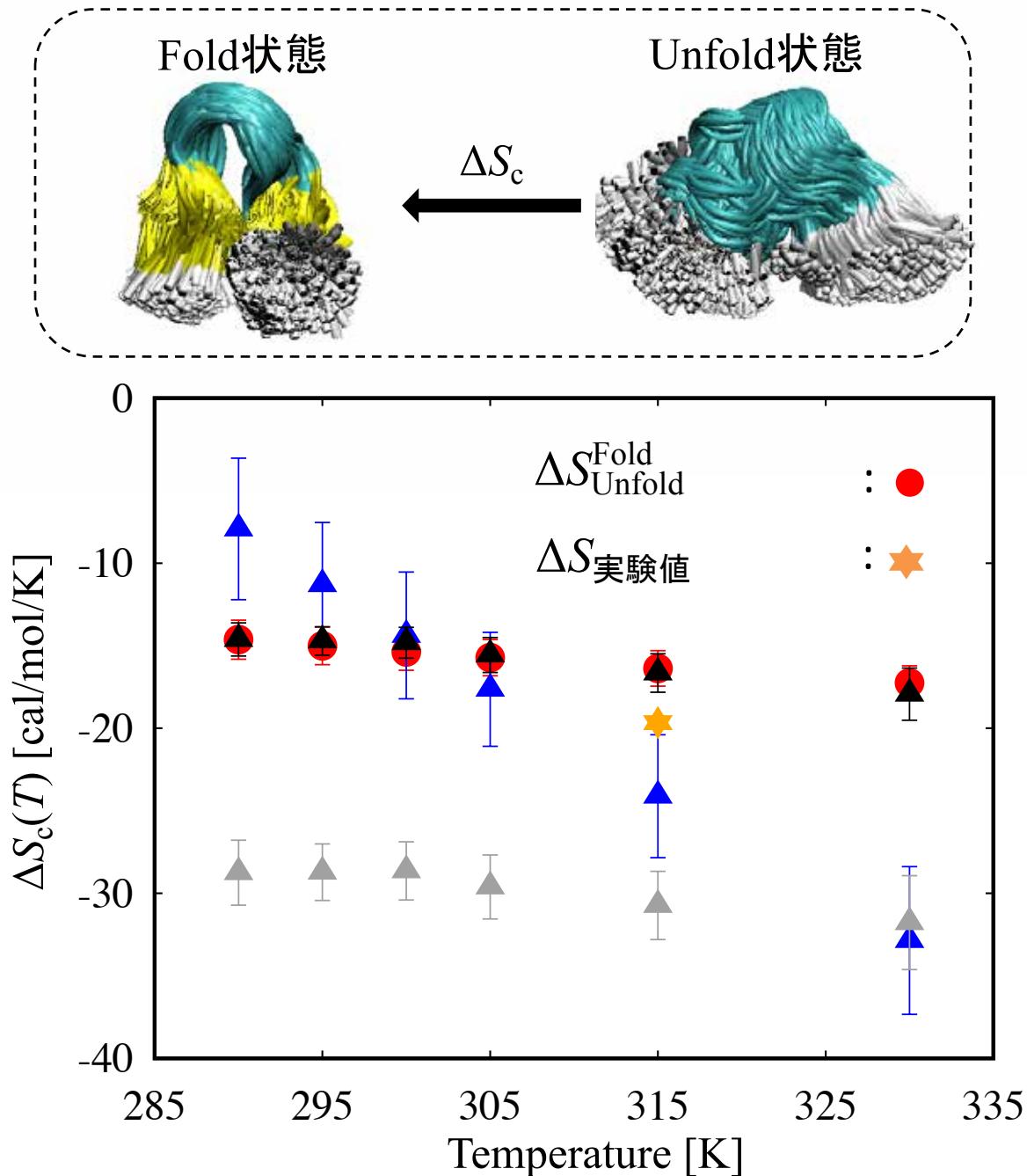


シニヨリンのFolding/Unfoldingのレプリカ交換MD(REMD)



- REMDでシニヨリンについて複数回のFolding/Unfoldingの転移を計算
- Folding/Unfoldingの転移曲線を得ることに成功
- REMDでの転移温度($T_m=305\text{K}$)は実験値(315K)に近い値

シニヨリンのFoldingに伴う $\Delta S_c(T)$ の計算



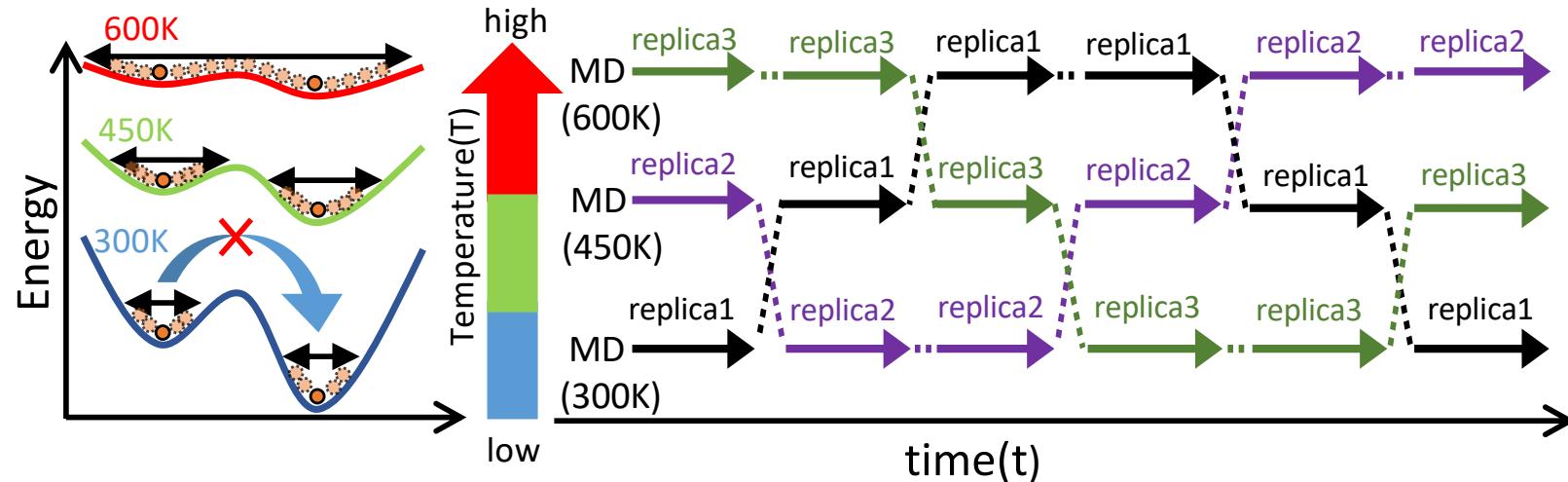
MDで解析できること(例)

- 立体構造のゆらぎを観る
 - 蛋白質のどこが柔らかく、どこが堅いか？
- 運動の相関を観る
 - 蛋白質のどこが連動して運動しているか？
- 相互作用を観る
 - リガンドとの相互作用は、どこが強いか？
- 熱力学量を得る
 - 結合自由エネルギー、エンタルピー、エントロピーを得る
- リガンド結合ポーズを得る
 - リガンドの安定な結合ポーズ予測

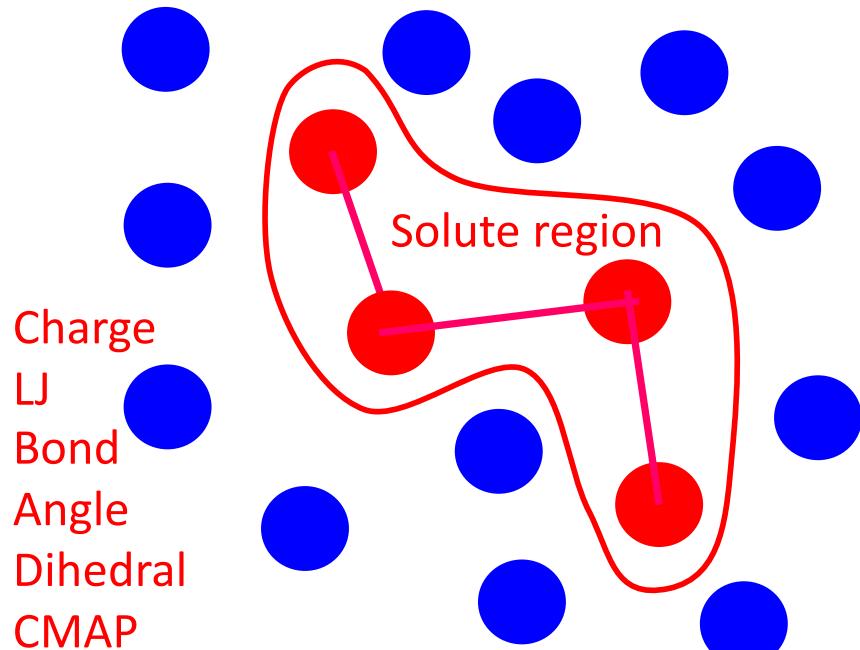
(Generalized) Replica Exchange with Solute Tempering (REST / gREST)

REST: Liu et al. PNAS, 102, 13749 (2005).

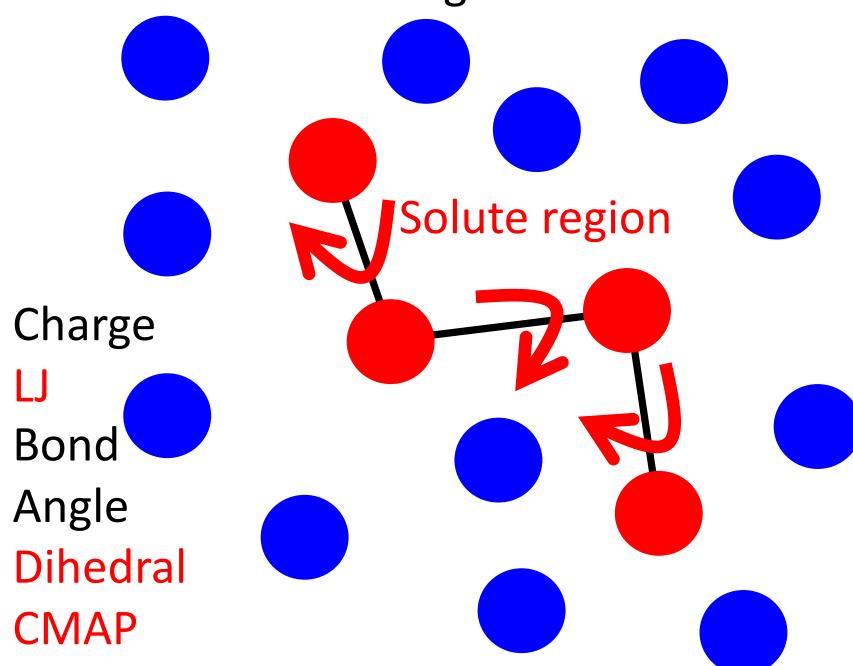
gREST: M.Kamiya and Y.Sugita, JCP, 149, 072304 (2018).



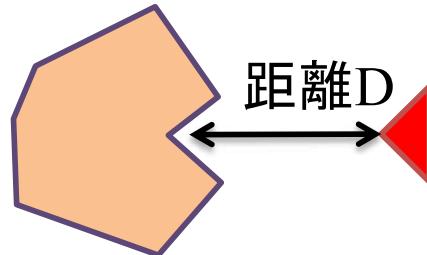
Conventional REST



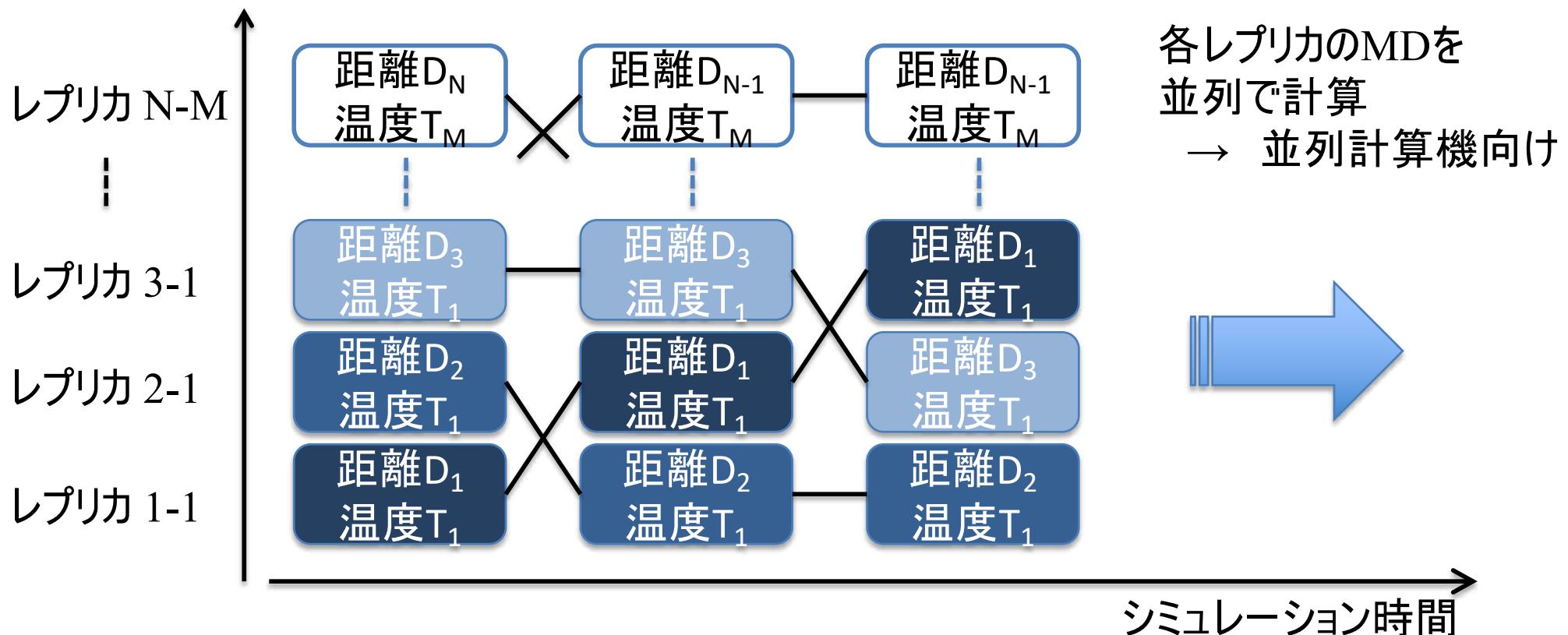
gREST



結合ポーズ推定



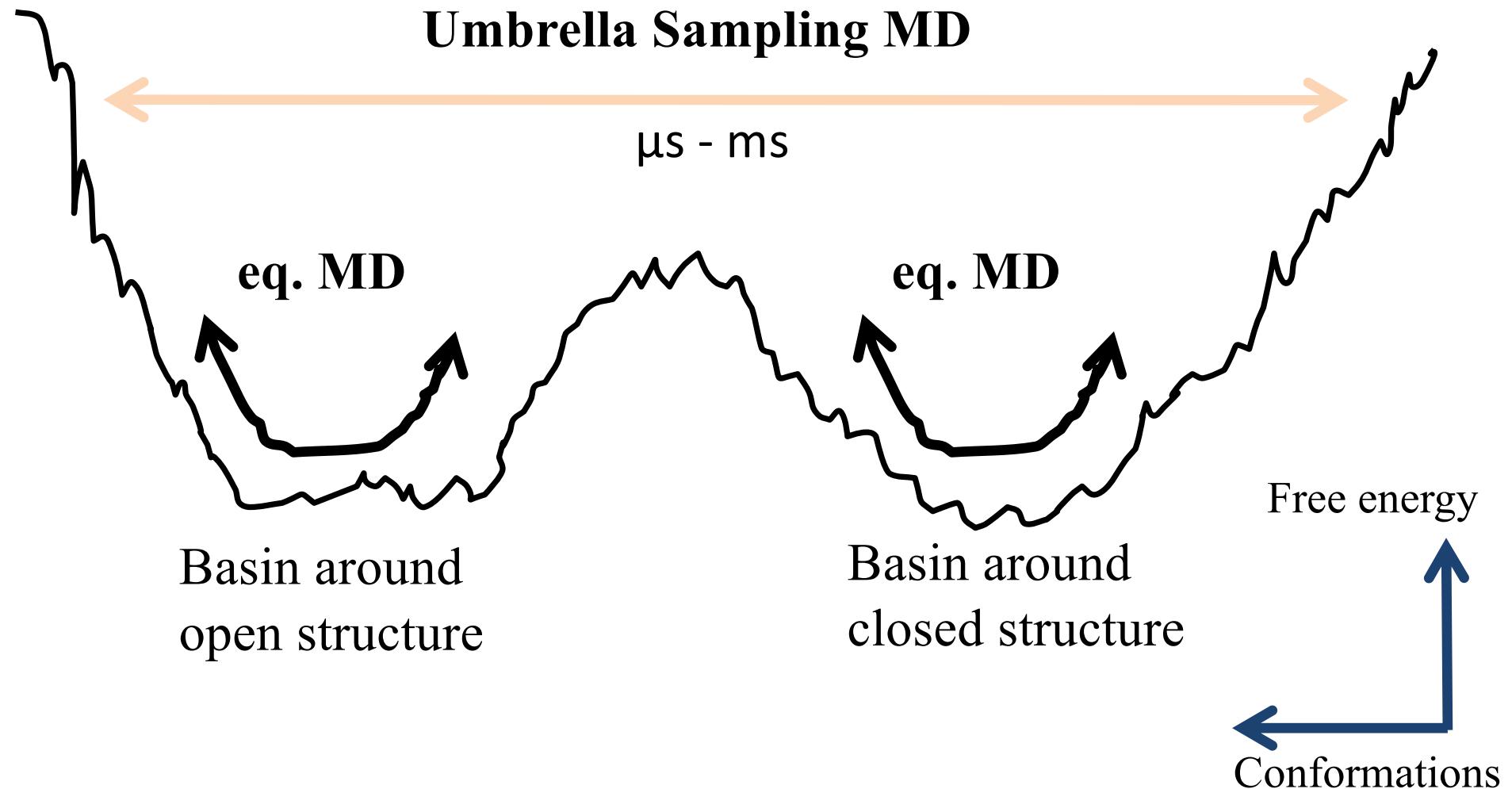
- ✓ レプリカ交換法によって結合距離とリガンド温度を変えながら、安定な結合状態を探索
- ✓ 2次元空間での
- ✓ 結合ポーズを推定可能



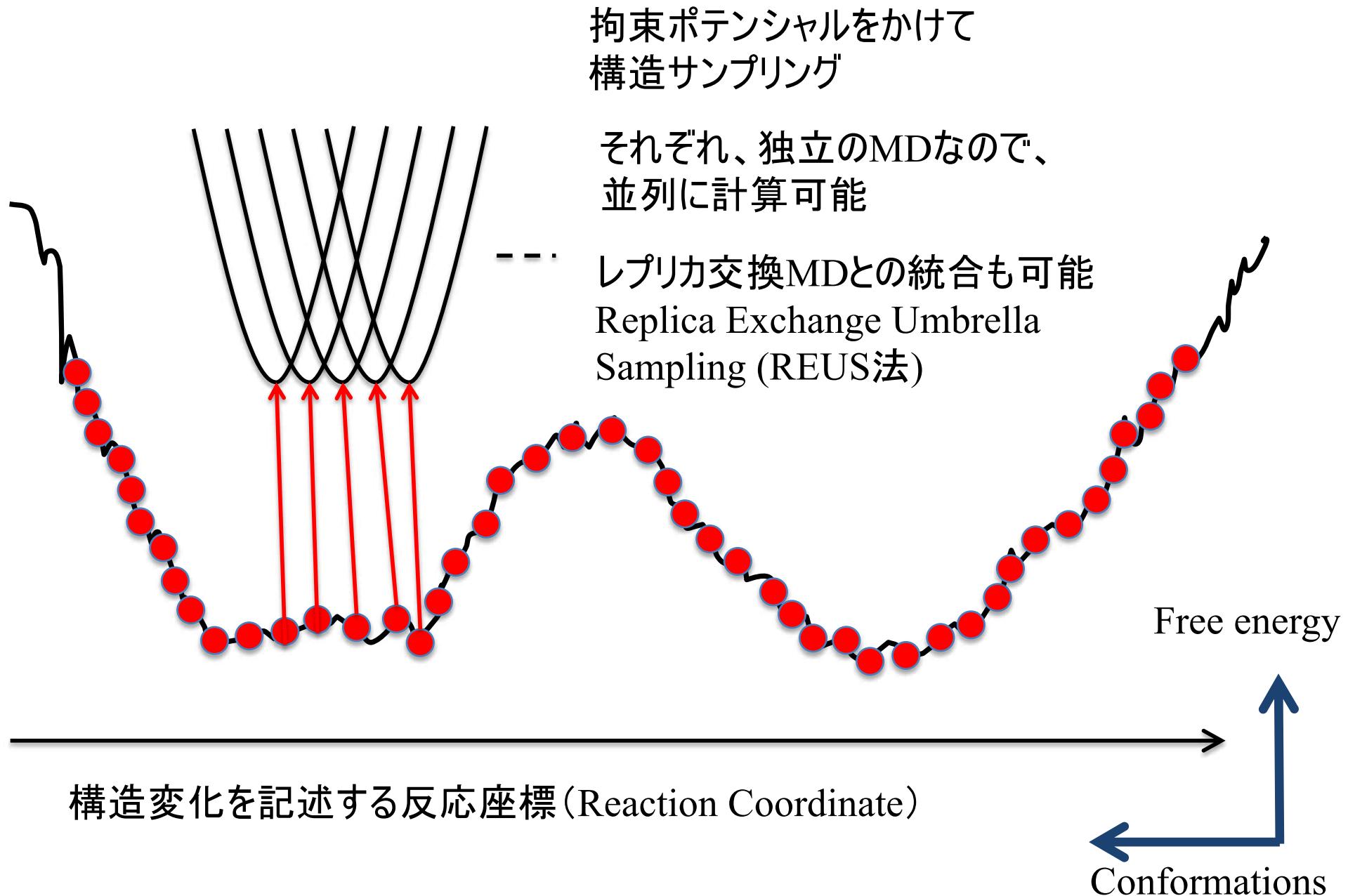
MDで解析できること(例)

- 立体構造のゆらぎを観る
 - 蛋白質のどこが柔らかく、どこが堅いか？
- 運動の相関を観る
 - 蛋白質のどこが連動して運動しているか？
- 相互作用を観る
 - リガンドとの相互作用は、どこが強いか？
- 熱力学量を得る
 - 結合自由エネルギー、エンタルピー、エントロピーを得る
- リガンド結合ポーズを得る
 - リガンドの安定な結合ポーズ予測
- 立体構造変化を観る
 - どのように、蛋白質は立体構造変化するのか？

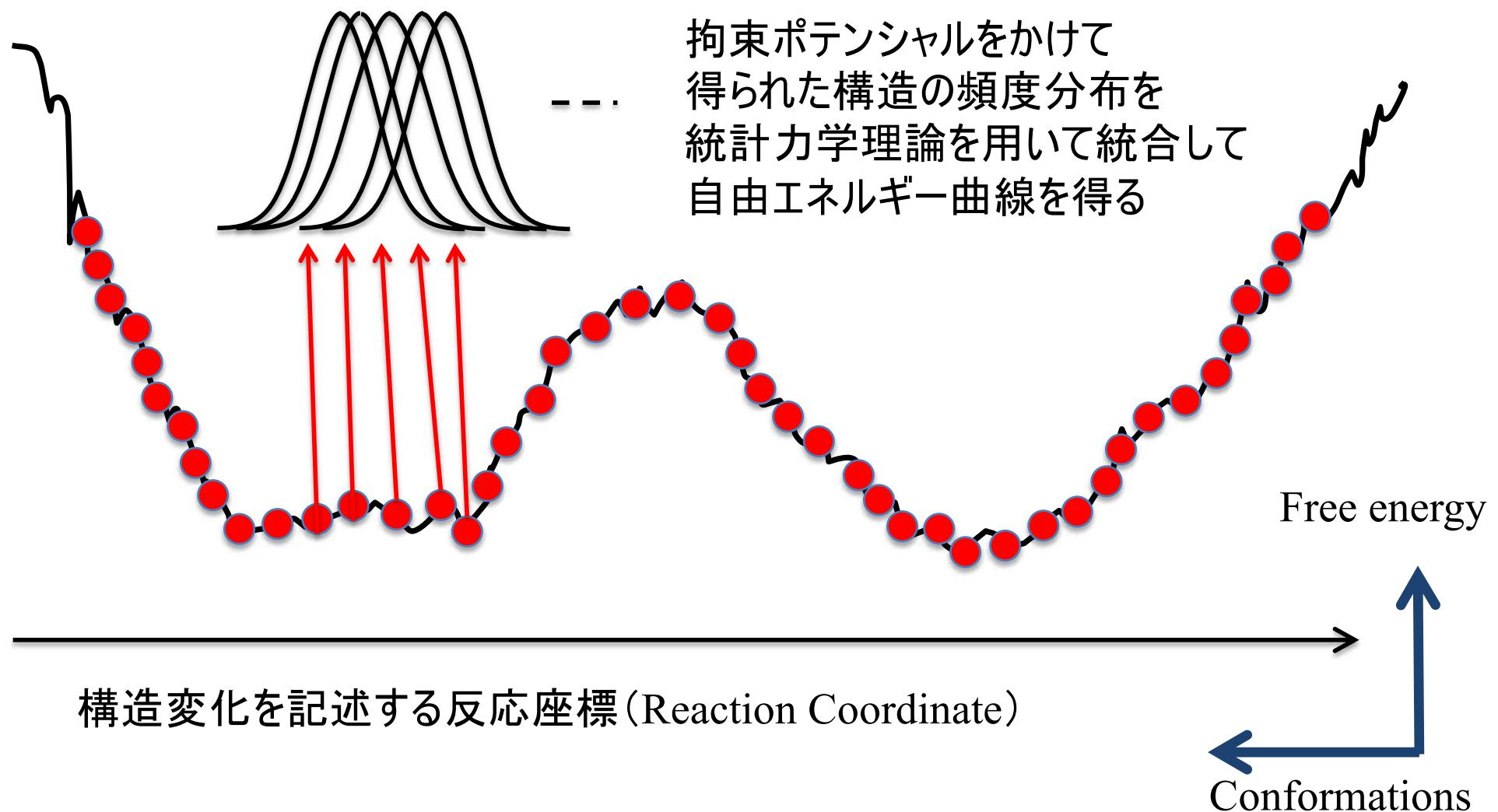
Method: umbrella sampling simulations



Method: umbrella sampling simulations



Method: umbrella sampling simulations

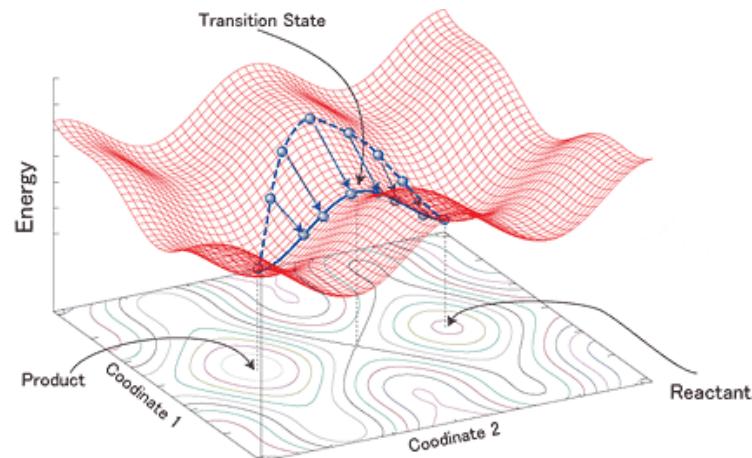


構造変化に対するFree Energy Simulations

1. Initial Path

Nudged Elastic Band

open close



2. Umbrella Sampling Simulations

For each intermediates along path, simulations with restraint,

$$w_j = K_{\text{rmsd}} (\Delta D_{\text{rmsd}} - \Delta D_{\min})^2$$

$$\Delta D_{\text{rmsd}} = \text{rmsd}(X, X_{\text{open}}) - \text{rmsd}(X, X_{\text{closed}})$$

was carried out.

Restraint potential is applied to both main chains and side chains.

B. Roux JACS (2005) C.L. Brooks PNAS (2007)

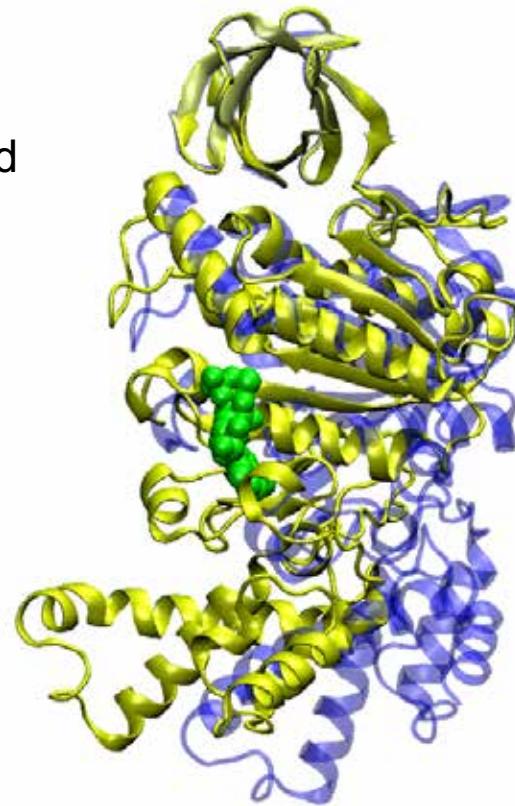
3. WHAM

Weighted Histogram

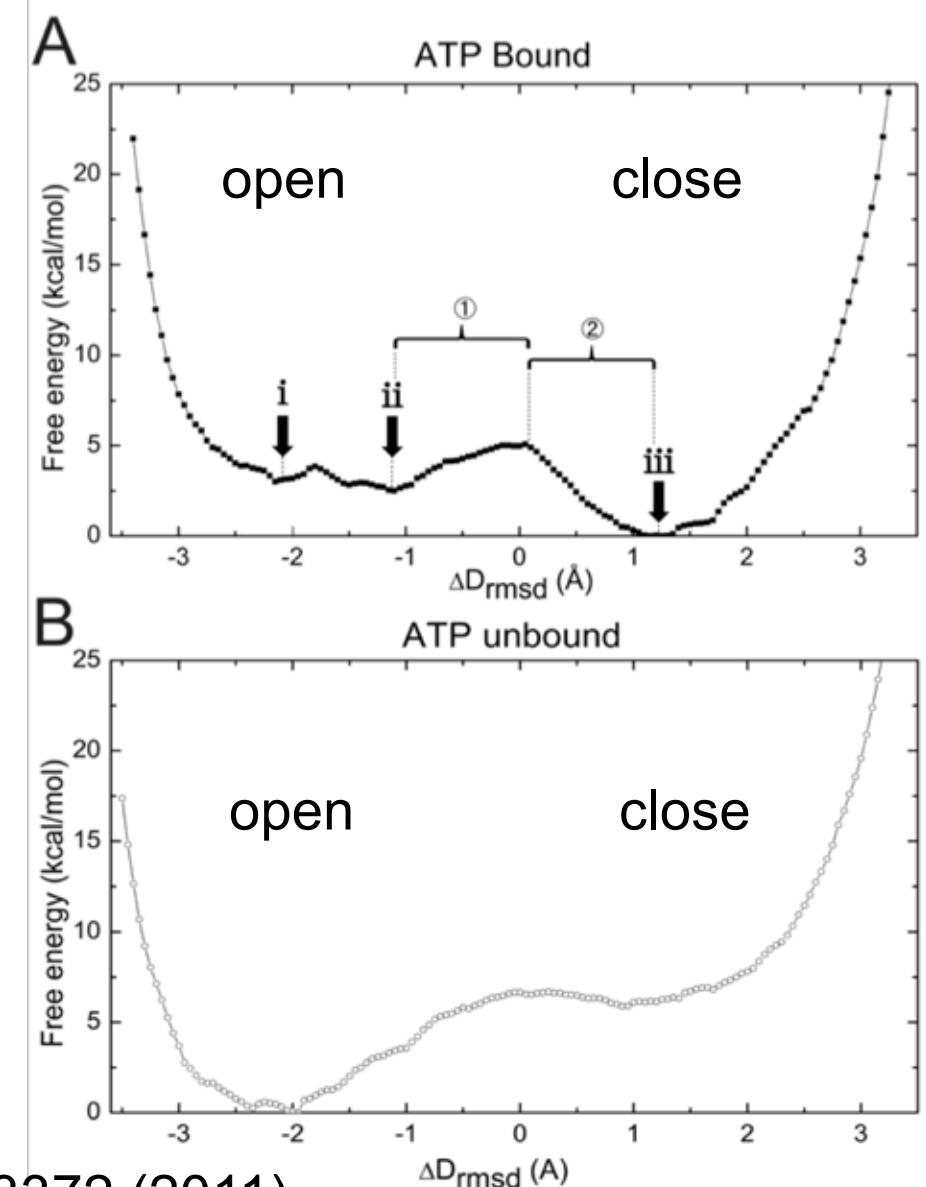
Analysis Method was used for removing restraints and calculating free energy profiles along structural transition between open and closed conformations.

Free Energy Profiles for Open-Close Transition

blue: open
yellow: closed
green: ATP

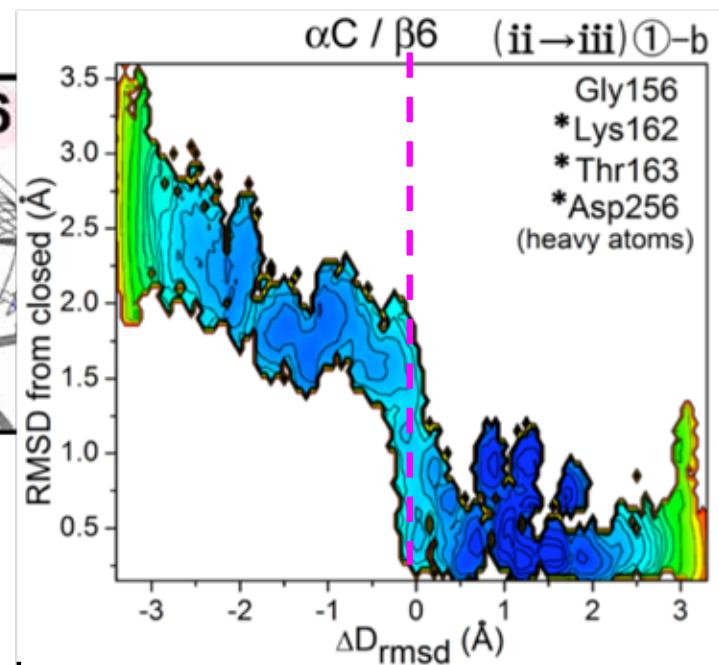
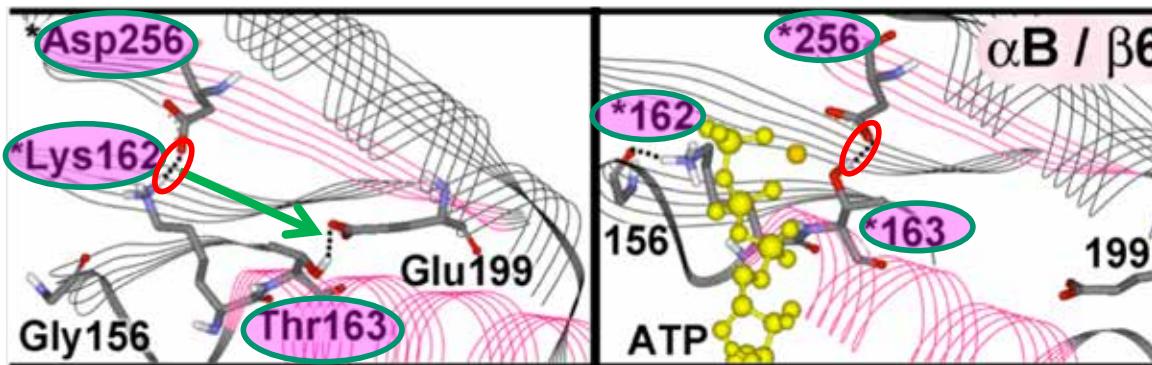
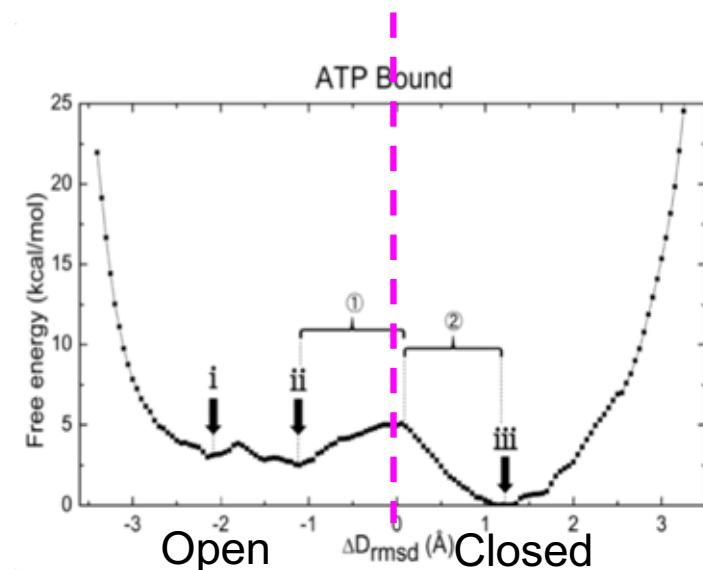
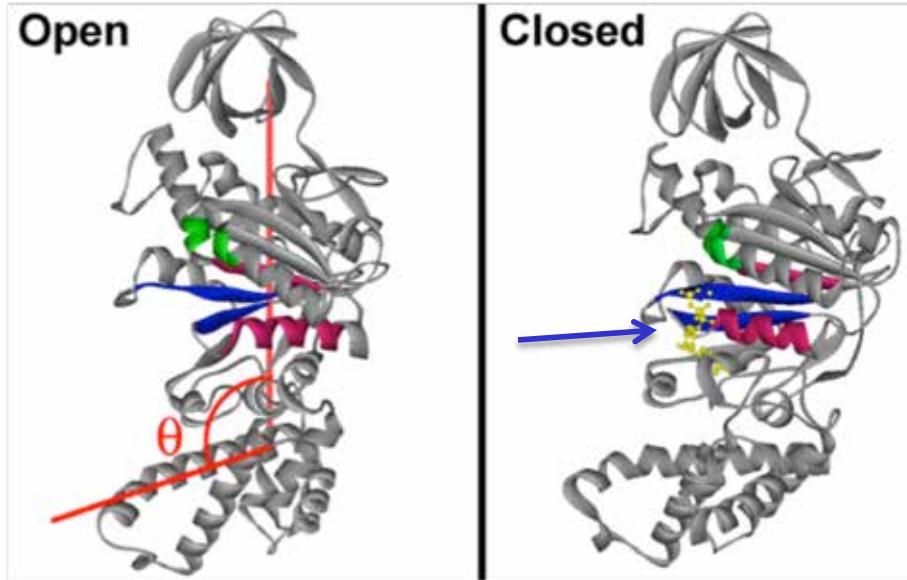


- ✓ Meta-stable state is found:
ATP-bound open conformation.
→ encounter complex



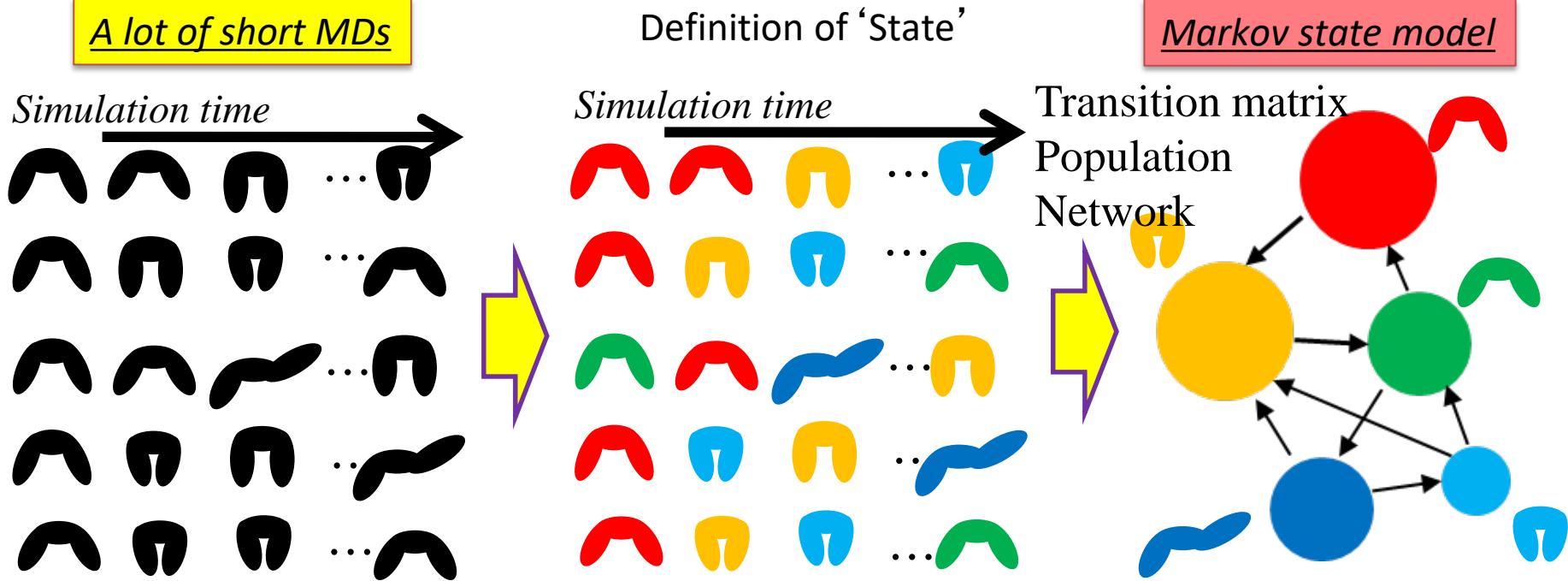
Ito, Oroguchi, Ikeguchi, JACS, 133, 3372 (2011)

ii → *iii* ① transition: H-bond in P-loop



H-bond partner of Asp256 is changed from Lys162 to Thr163 (P-loop). Mutation of Asp256, Lys162, or Thr163 results in remaining open conformation even with ATP bound.

マルコフ状態遷移モデルによるダイナミクス解析



全原子-粗視化連成MDシミュレーション

All atom (MM)

Slow motion
High precision

Coupling force

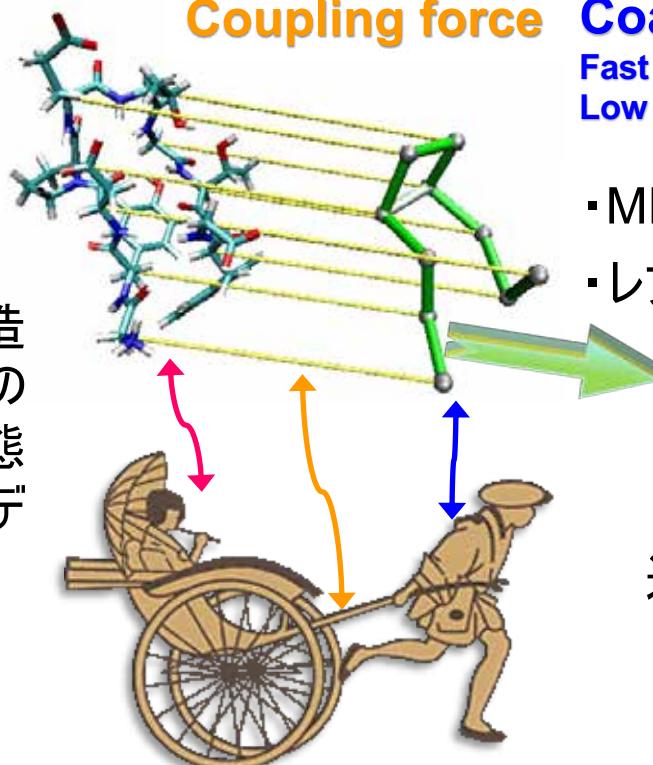
Coarse-grained (CG)

Fast motion
Low precision

森次、木寺ら

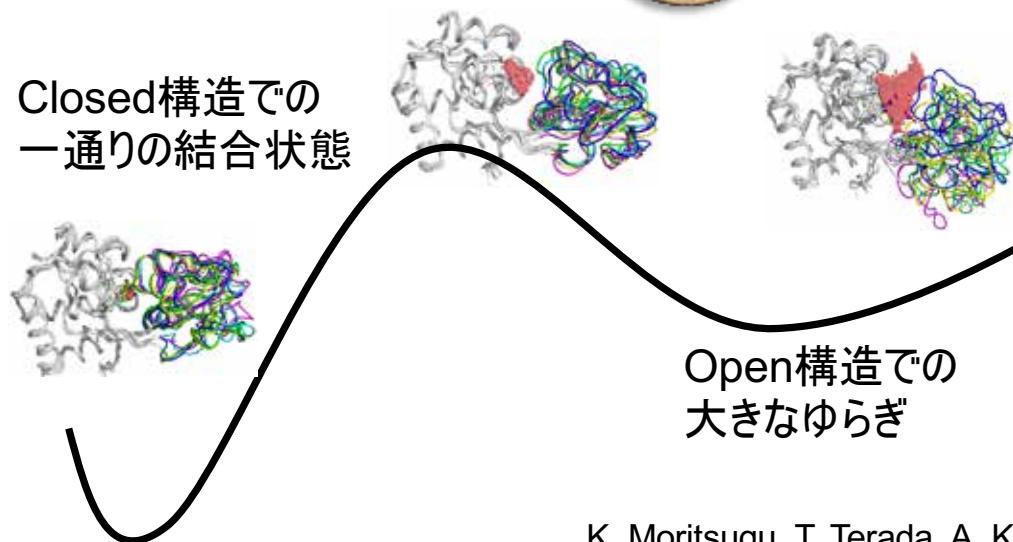
- ・MMをCGと連成させた広範なサンプリング
- ・レプリカ交換法によりMMだけの情報を得る

目的：大規模な構造変化をするタンパク質の阻害剤を非結合状態の結晶構造のみからデザインする方法の確立



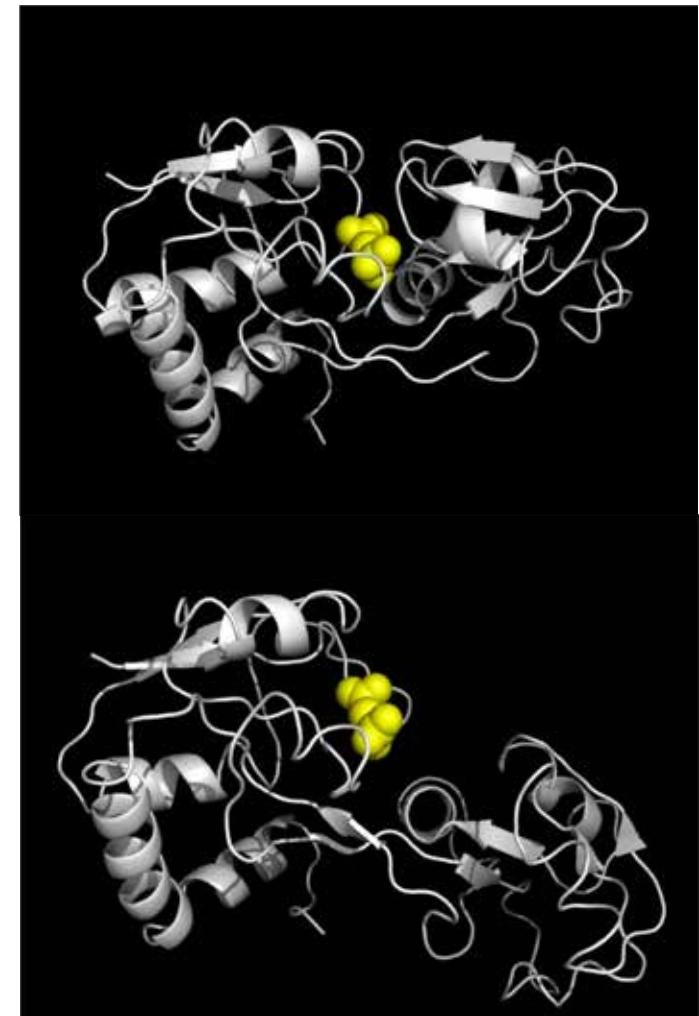
通常MD

Closed構造での一通りの結合状態



Open構造での大きなゆらぎ

MM-CG
連成MD



グルコキナーゼの全原子構造探索による活性化機構の解析

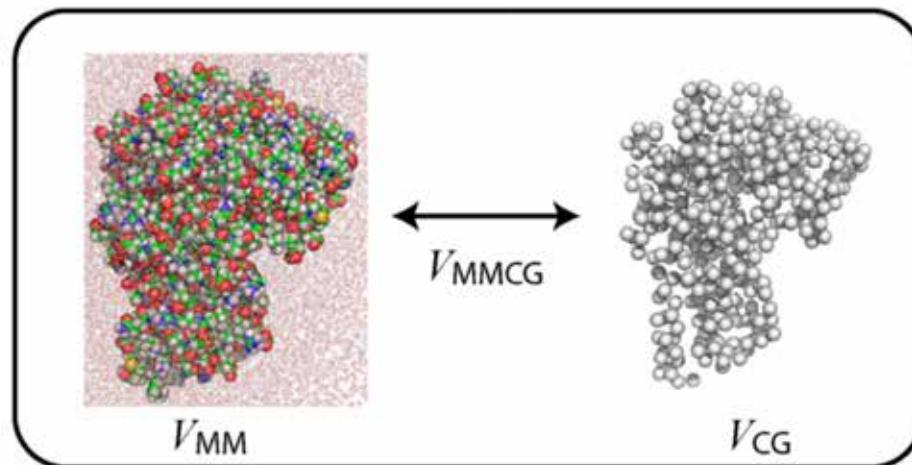
糖尿病などの創薬ターゲットである
グルコキナーゼ活性化の
構造ダイナミクス基盤



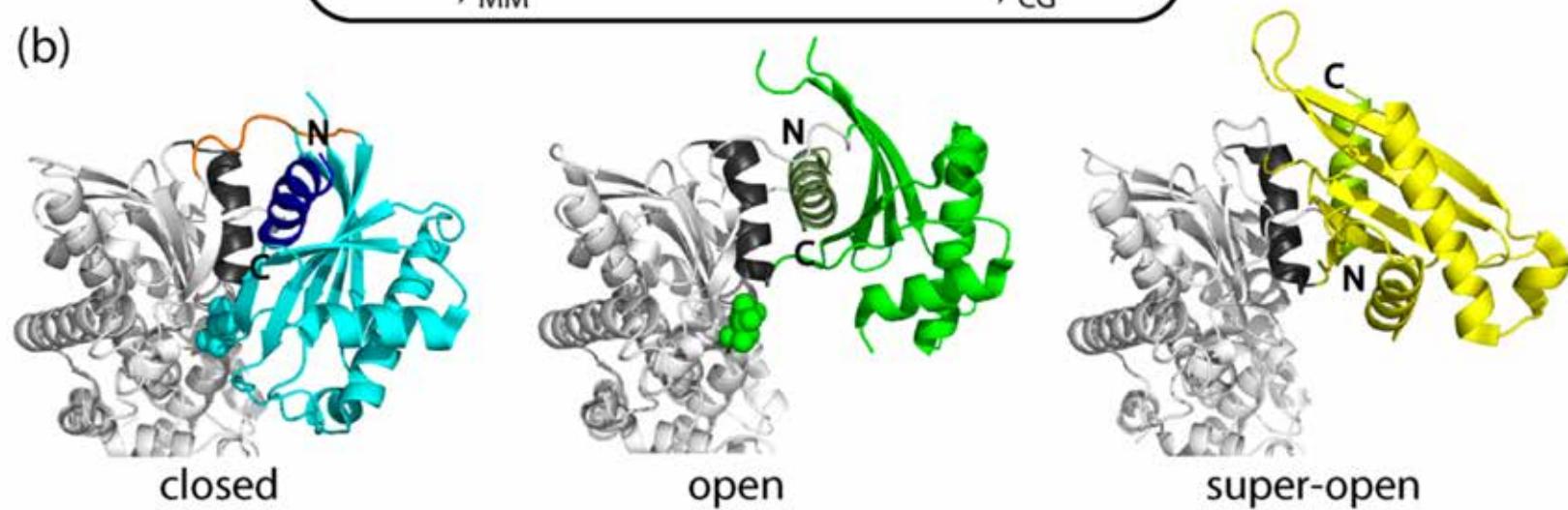
MSES法による全原子構造探索と
weighted ensemble法の組み合わせ

Moritsugu et al., JCP 2018

(a)



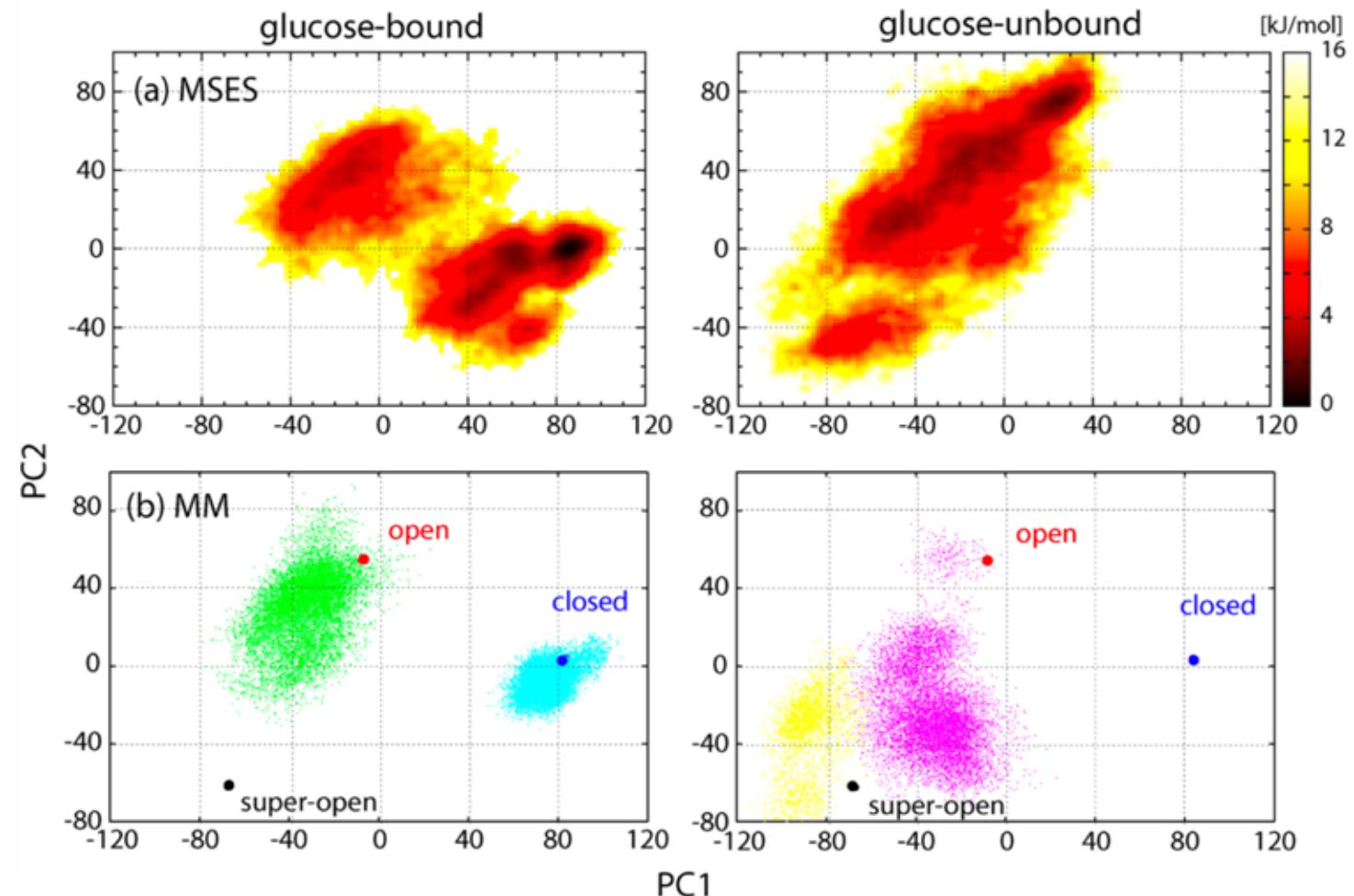
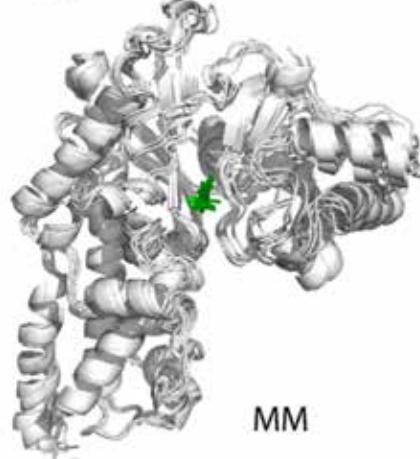
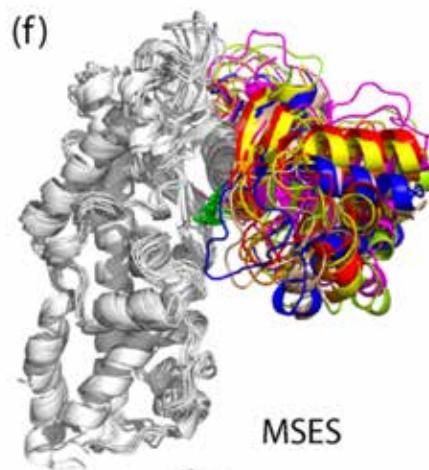
(b)



グルコキナーゼの全原子構造探索による活性化機構の解析

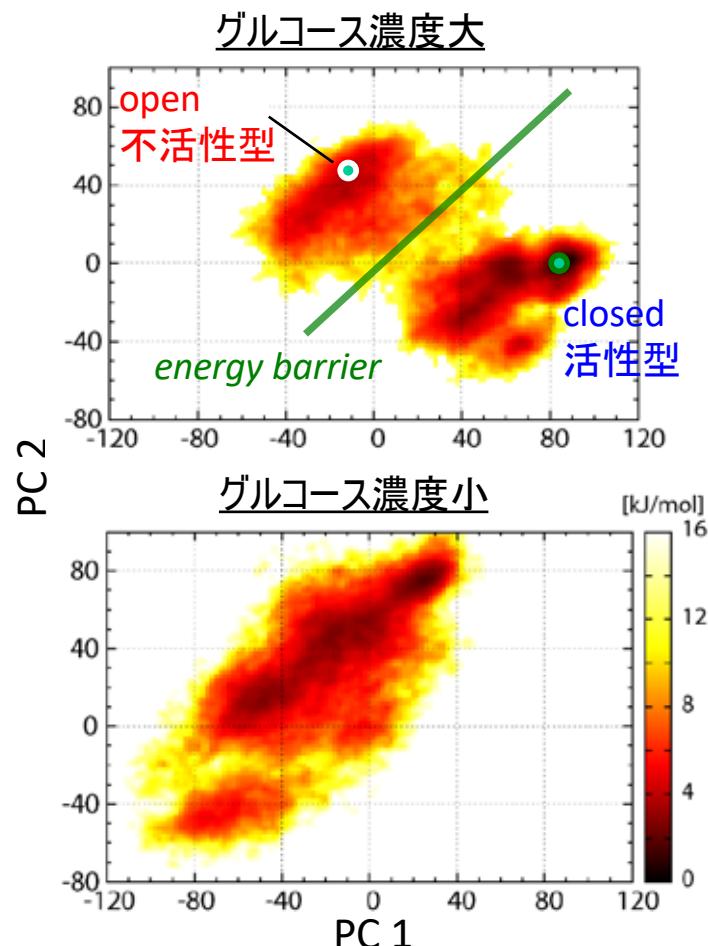
糖尿病などの創薬ターゲットである
グルコキナーゼ活性化の
構造ダイナミクス基盤

MSES法による全原子構造探索と
weighted ensemble法の組み合わせ



グルコキナーゼの全原子構造探索による活性化機構の解析

糖尿病などの創薬ターゲットである
グルコキナーゼ活性化の
構造ダイナミクス基盤

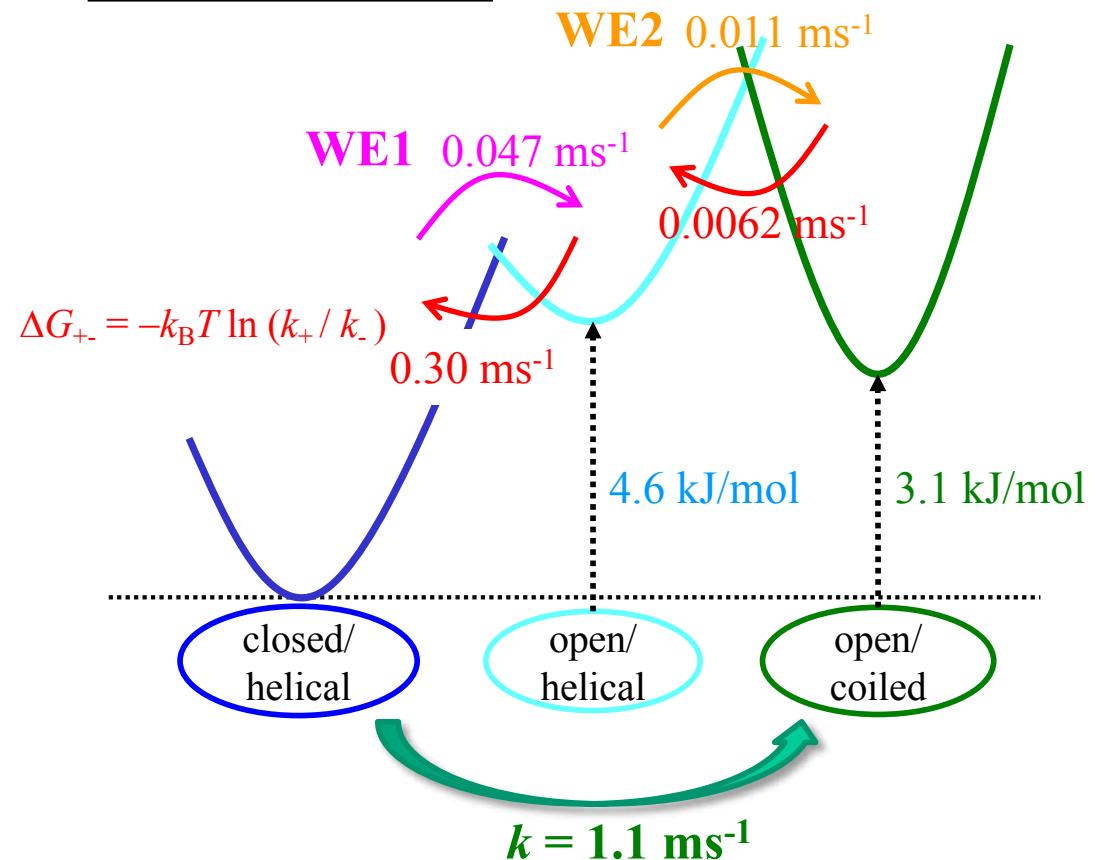


グルコース濃度によるconformational switch



MSES法による全原子構造探索と
weighted ensemble法の組み合わせ

3-basin kinetic model:



ミリ秒オーダーの構造変化時定数を計算

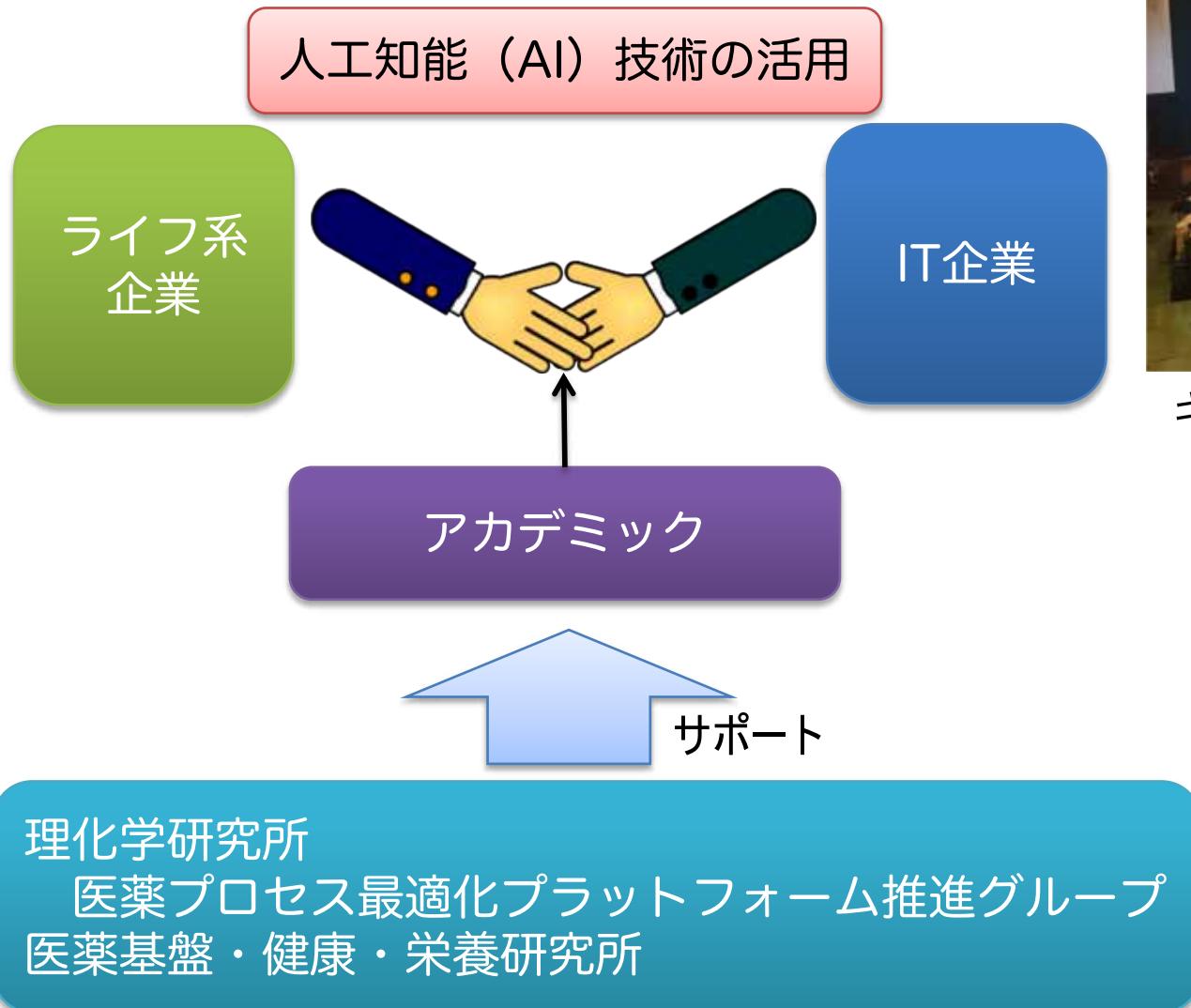
MDで解析できること(例)

- 立体構造のゆらぎを観る
 - 蛋白質のどこが柔らかく、どこが堅いか？
- 運動の相関を観る
 - 蛋白質のどこが連動して運動しているか？
- 相互作用を観る
 - リガンドとの相互作用は、どこが強いか？
- 熱力学量を得る
 - 結合自由エネルギー、エンタルピー、エントロピーを得る
- リガンド結合ポーズを得る
 - リガンドの安定な結合ポーズ予測
- 立体構造変化を観る
 - どのように、蛋白質は立体構造変化するのか？
- 実験結果と一致する構造を得る
 - NMRやSAXSなどの実験結果と整合性を検討

Life Intelligence Consortium(LINC)

ライフ インテリジェンス コンソーシアム
(Life Intelligence Consortium ; LINC)

89組織(製薬企業、IT企業、アカデミア)から構成



キックオフミーティングの様子

10のワークグループ

1. 未病・先制医療
2. 臨床・診断
3. 創薬テーマ創出
4. 分子シミュレーション
5. メドケム・分子設計・ADMET
6. トランスレーショナル・リサーチ
7. バイオロジクス・製剤・ロボティクス
8. 治験・市販後・メディカル・アフェアーズ
9. 知識ベース・NLP
10. AI基盤



池口満徳
WG4代表



病気の一歩手前

予防・先制医療

WG1

未病・先制医療

- ・健康診断データによる発症予測
- ・マイクロバイオーム、オミクスデータ解析
- ・デジタルヘルス



WG3

創薬テーマ創出

- ・有望提携先や研究テーマの自動探索
- ・標的分子探索
- ・ドラッグリポジショニング

WG6

トランスレーショナルリサーチ

- ・非臨床データからのヒトADMET予測
- ・疾患メカニズム解明・ブリッジング予測

WG7

バイオロジクス・製剤・ロボティクス

- ・バイオロジクス関連AI
- ・結晶形・製剤関連AI
- ・調剤ロボティクス

WG8

治験・市販後・メディカルアフェアーズ

- ・AIによる治験の効率化
- ・有害事象の情報基盤
- ・製品Q&Aシステム
- ・アウトカムリサーチ・医療技術評価

WG9

知識ベース・自然言語処理

知識ベース/計算機サーバー

WG10

AI基盤

- ・ライフサイエンスのためのAI基盤

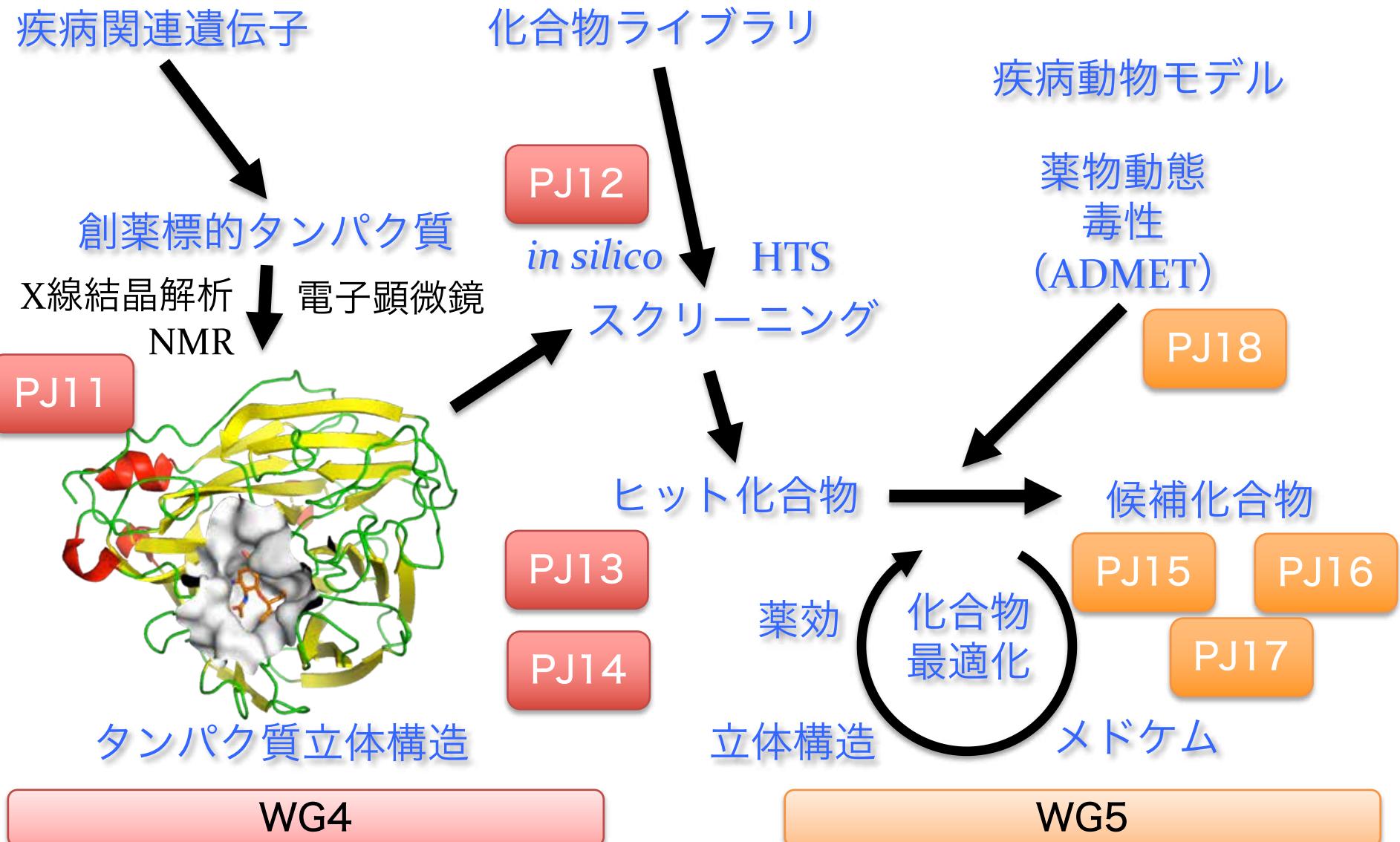


本間光貴
WG5代表

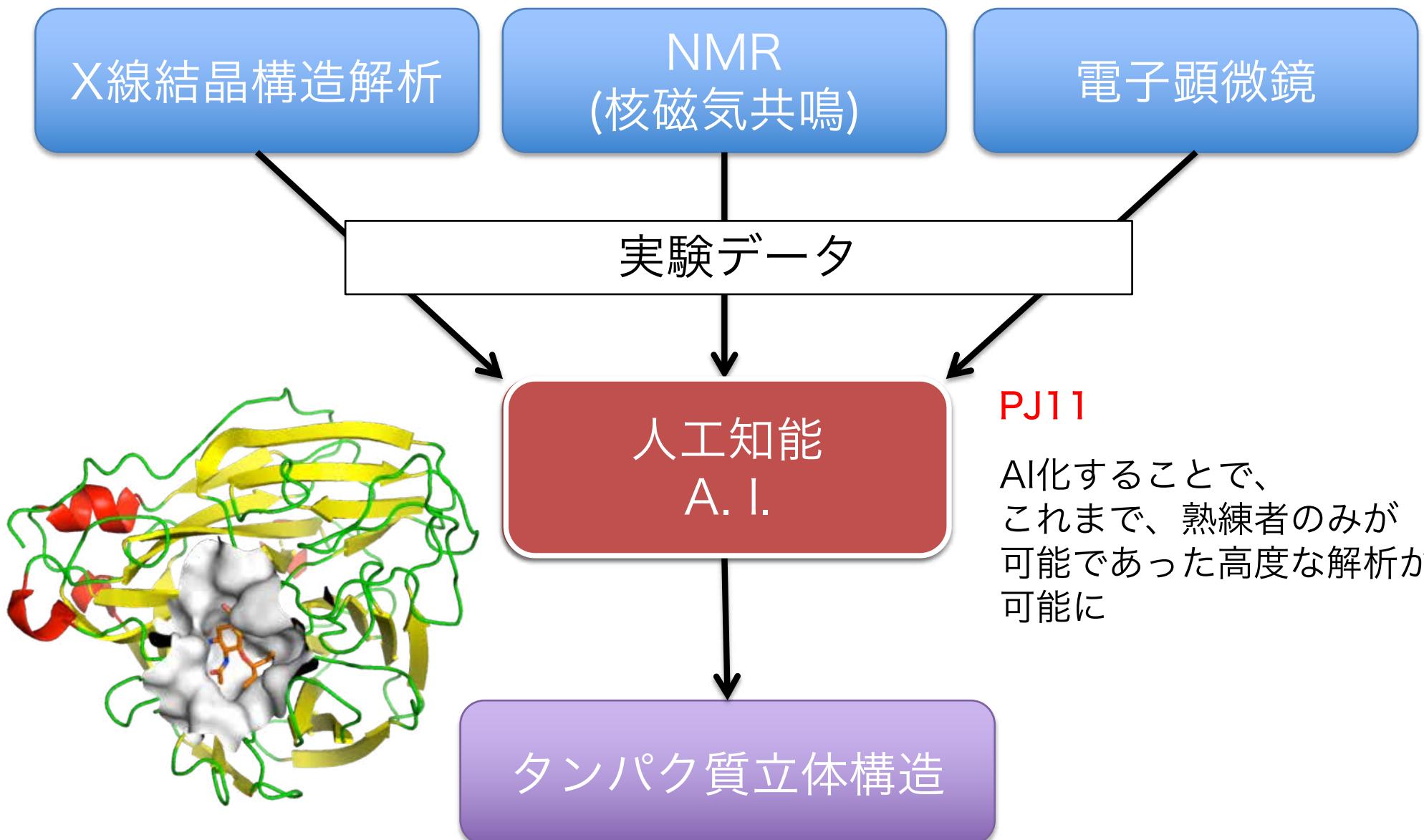


奥野恭史
LINC代表

LINCパンフレットより

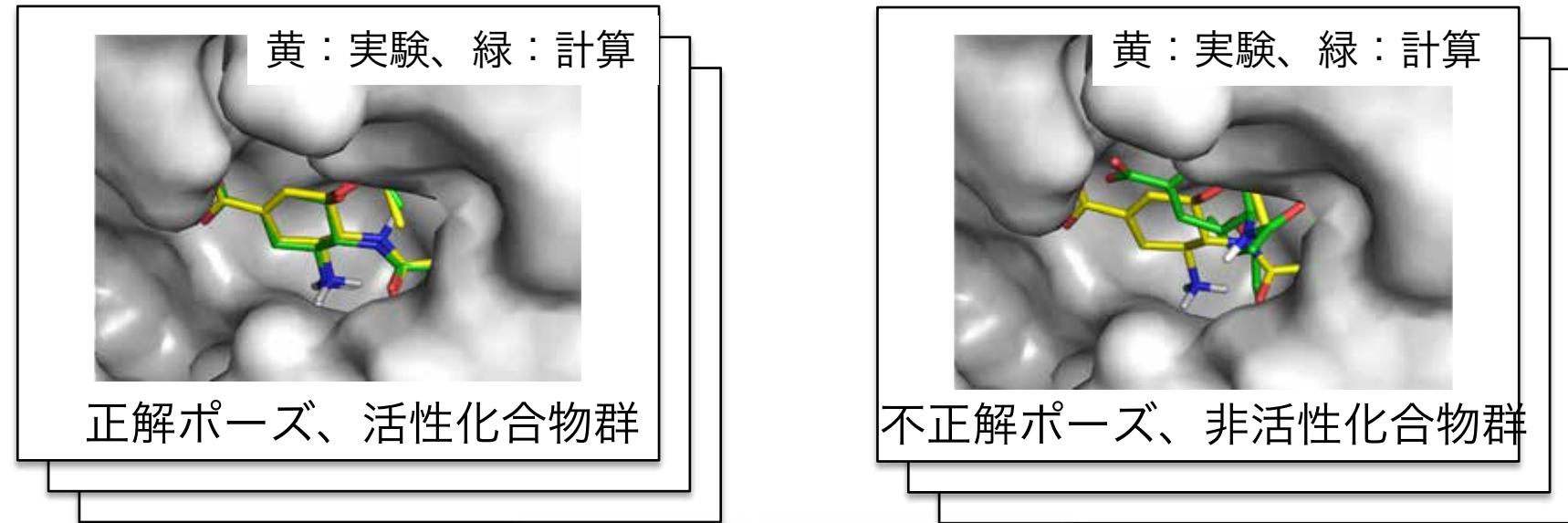


立体構造決定法

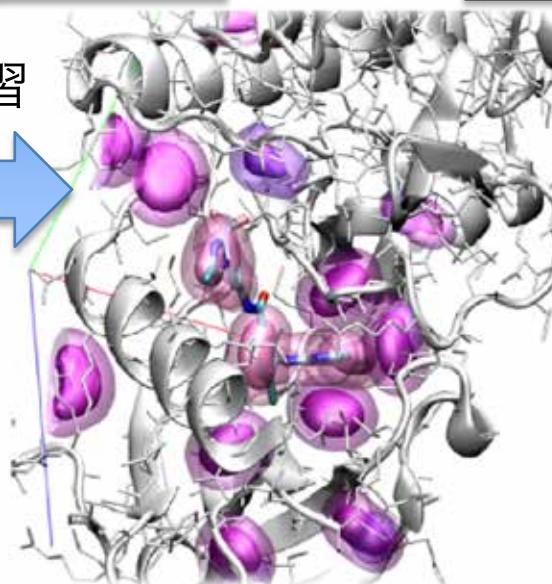


PJ12：AIによるドッキング計算高度化

これまでの問題点：精度の低さ（1~2%のヒット率）

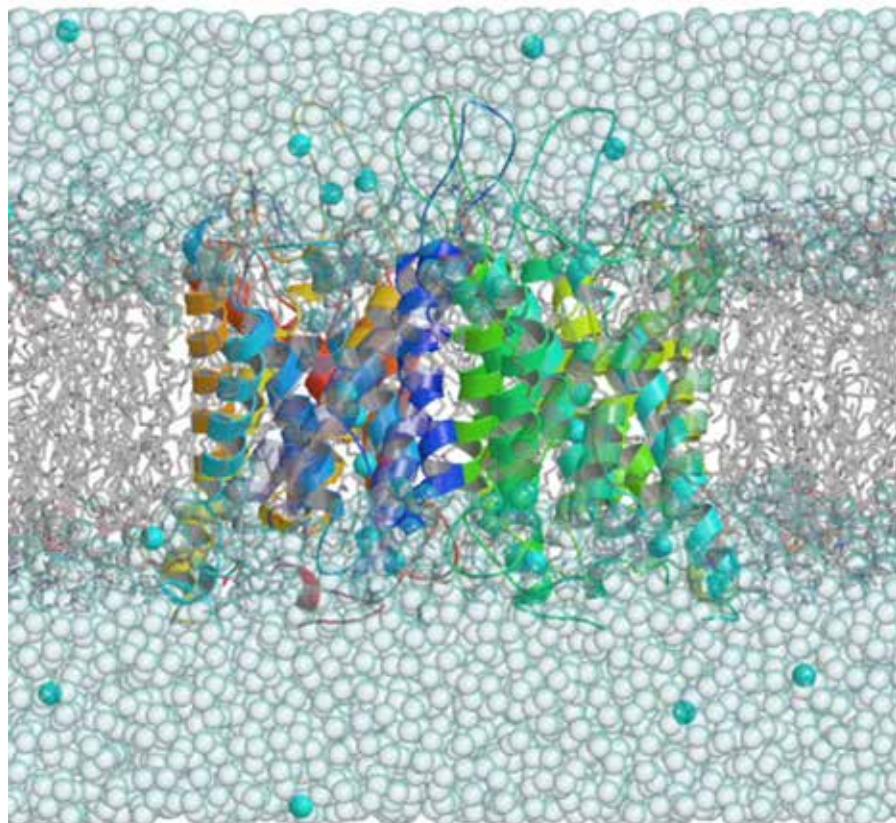


AIは、画像認識のように、正しい結合を間違った結合から見出すことができるか？



画像認識	結合計算
2次元	3次元
色 (RGB)	原子種
対象 (猫等)	結合ポーズ

PJ13：分子動力学計算におけるAI活用



膜タンパク質アクアポリンのMD
全原子数～16万
数千万～数十億ステップの計算

これまでの問題点

時間軸を含む大量のデータ
(静止画像 → 動画)

熟練した技術者が長時間かけて解析



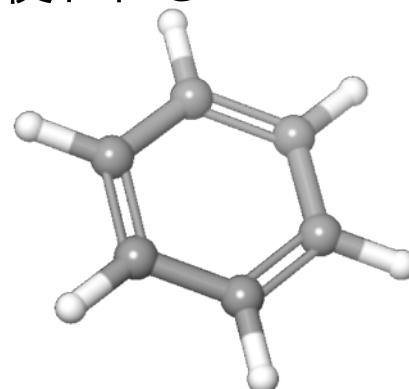
AIによって、重要な箇所を発見

動的なタンパク質の制御による
創薬のブレークスルーを目指す

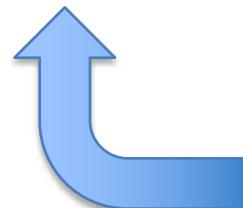
PJ14：AIを用いた高精度分子力場

分子力場

MDでよく使われる



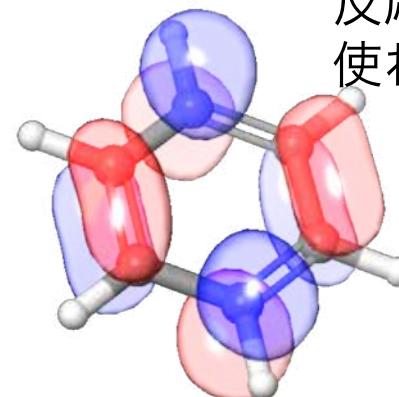
分子運動	電子状態
○	X



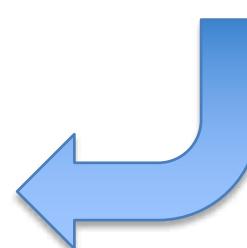
電子状態を加味した
高精度力場を高速計算

分子軌道

反応解析でよく
使われる



分子運動	電子状態
X	○



電子状態を学習

人工知能
A.I.

分子運動	電子状態
○	○

まとめ

- ・ 分子動力学シミュレーションの新展開
 - スパコン「京」から「富岳」へ
 - アンサンブル シミュレーションの有効性
- ・ 分子動力学シミュレーションの解析・応用
 - 分子動力学シミュレーションの活用例
- ・ 分子シミュレーションと人工知能(AI)の連携
 - LINCコンソーシアム：WG4
 - 分子子ミューレヨン分野でのAIの活用